Allelic imbalance and somatic mutations in folate pathway genes in childhood acute lymphoblastic leukaemia

GEOFFREY BOSSON

MSc CertHSM CertTQM CertT(HE) CSci FIBMS

A thesis submitted in partial fulfillment of the requirements of the
University of Northumbria at Newcastle
for the degree of
Doctor of Philosophy

Research undertaken in the
School of Applied Sciences, University of Northumbria at Newcastle
and
Northern Institute of Cancer Research, Newcastle University Medical School

May 2008

ABSTRACT

Acute lymphoblastic leukaemia (ALL) is the commonest of the childhood cancers, but thankfully responds well to chemotherapeutic agents, with 80% of children achieving long-term survival. However, for the remaining 20% who relapse, outcome is bleak. Increasing knowledge and understanding of pharmacogenetics indicates that constitutive or acquired resistance to the drugs used in the treatment of cancer contribute to relapse. Methotrexate (MTX) is one of the most important drugs used in the treatment of ALL and the work presented here contributes to the body of knowledge relevant to the understanding of resistance to this drug. The aim of this research is to determine if changes in genes involved in folate metabolism contribute to methotrexate resistance and subsequent relapse of childhood ALL.

The reduced folate carrier (RFC) is required to transport methotrexate into the cell where it competitively inhibits dihydrofolate reductase (DHFR) and other key enzymes of folate and 1-carbon metabolism. Following the design of primers and optimisation of PCR amplification, the entire coding regions of both RFC and DHFR were screened for novel SNPs or mutations using denaturing high performance liquid chromatography (DHPLC) on genomic DNA from 40 normal and 40 relapse childhood ALL patients. Ethical approval for use of the samples was granted under WCRLEC no347 and reference 2002/111 for normal and relapse samples respectively. The relapse group was made up of 29 males and 11 females with an average age of 6.59 years (range 0.8-14.1) and was made up of 5 T-cell; 30 B-cell; 4 mixed lineage; and 1 null classification. The screening method was shown to be sufficiently sensitive to detect single base changes. Several of the exons in both genes have a high G-C content and required destabilisation agents and/or use of a new DNA polymerase (Optimase™) to achieve sufficient PCR amplification. Dimethyl sulphoxide (DMSO) and betaine were shown to be effective agents which did not interfere with subsequent DHPLC.

The results show that somatic mutations in the coding sequences of RFC and DHFR are rare in relapsed ALL and that while it is recognised as a mechanism of methotrexate resistance *in vitro*, it is unlikely to contribute to relapse in children with ALL. However, for one patient who suffered multiple relapses, a novel acquired mutation was identified in the 5'-UTR of the RFC-1 gene (C-37T). The significance

of the C-37T mutation on RFC transcription requires further study, but it may decrease RFC mRNA quantity or stability and thus protein levels.

DHPLC analysis also detected common SNPs. In terms of frequency, there were no significant differences between relapse and normal samples for the genotypes; RFC G80A (38.3% G/G; 49.4% G/A; 12.3% A/A in 81 normal samples; 31.8%; 56.8%; 11.4% respectively in 44 relapse samples; χ^2 p = 0.656); RFC C696T (97.8% C/C; 2.2% C/T; 0% T/T in 45 normal samples; 92.5%; 5%; 2.5% respectively in 40 relapse samples; χ^2 p = 0.458). However, there was a significant difference in the 5'-UTR -19 base pair deletion in the DHFR gene between the normal and relapse groups (0% WT/WT; 100% WT/-19 or -19/-19 in 45 normals; 26.7% and 73.3% respectively in 15 relapse samples; Fisher's exact probability p = 0.017) and needs to be confirmed in a larger cohort.

Estimated copy number and the probability of loss of heterozygosity (LOH) were generated from the Affymetrix 50K SNP microarray for the RFC, DHFR and other genes involved in folate metabolism using gDNA from 73 presentation and 20 relapse childhood ALL cases. Mann-Whitney non-parametric comparison of the presentation and relapse data showed that for estimated copy number, statistically significant differences were seen for MTHFR (p = 0.0072), MS (0.0025), FPGS (0.00048), TS (0.00046), CBS (0.0002) and RFC (0.0001). When the data at the nearest SNP location to the gene was presented as a scatterplot, the difference in each case was due to a bimodal distribution of the presentation data with a subpopulation of samples with higher values which was not seen in the relapse group. Similar analysis of the LOH data did not show any p values less than 0.01, indicating that there were no differences between the two groups as a whole. However, the scatter plots of SNP data closest to the genes studied show individual samples which are of interest. For some samples the higher probability for LOH could be attributed to a decreased copy number verified by cytogenetic records, but others appear to have high probability for LOH with normal copy number which can result from uniparental disomy. The data provided here, does suggest that LOH in some patients may play a role in response to MTX and also supports previous findings that increased copy number of chromosome 21 is an indicator of good prognostic outcome.

LIST OF CONTENTS

2 SCREENING THE REDUCED FOLATE CARRIER

LIST OF APPENDICES

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A or a	Adenosine
ACN	Acetonitrile
AICART	Aminoimidazole carboxamide ribonucleotide transformylase
ALL	Acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
BMT	Betaine methyltransferase
bp	Base pair
C or c	Cytosine
1-C	One-carbon
CBS	Cystathionine β-synthase
CD	Clusters of differentiation
cDNA	Complementary deoxyribonucleic acids
CFU	Colony Forming Unit
CGH	Comparative genomic hybridization
CSF	Cerebrospinal fluid
DDATHF	$N^{5,10}$dideazatetra-hydrofolate
DHFR	Dihydrofolate reductase
DHPLC	Denaturing high performance liquid chromatography
DMSO	Dimethylsulphoxide
dNTPs	Deoxynucleotides
dsDNA	Double stranded DNA
dAMP	Deoxyadenosine monophosphate
dGMP	Guanosine monophosphate
dTMP	Deoxythymidylate monophosphate
dUMP	Deoxyuridine monophosphate
EMBL	European Molecular Biology Laboratory
FAB	French-American-British
FAD	Flavin adenine dinucleotide
FISH	Fluorescent *in-situ* hybridisation
5-formylTHF	N^5-formyltetrahydrofolate
5-FU	5-fluorouracil (5-FU)
FPGH	Folylpolyglutamyl hydrolase
FPGS	Folylpolyglutamate synthetase
G or g	Guanosine
γGH	Gammaglutamyl hydrolase
GART	Glycinamide ribonucleotide transformylase
Hcys	Homocysteine
^{3}H-MTX	Tritiated methotrexate
HPLC	High performance liquid chromatography
HSR	Homogenously staining regions
IFC-1	Intestinal folate carrier

IIR	Ion-interactive reagent (IIR)
K_M	Michaelis constant
LOH	Loss of heterozygosity
LRF	Leukaemia Research Fund
MDR	Multidrug resistance
5-methylTHF	N^5-methyltetrahydrofolate
MLL	Mixed Lineage Leukemia
6-MP	6-mercaptopurine
mRNA	Messenger ribonucleic acids
MRP	Multi-drug resistant protein family
MS	Methionine synthase
MTHFR	Methylenetetrahydofolate reductase
MTRR	Methionine synthase reductase
MTX	Methotrexate
MTX_{glu}	Polyglutamylated MTX
NAD(H)	Nicotinamide adenosine dinucleotide (reduced)
NADP(H)	Nicotinamide adenosine dinucleotide phosphate (reduced)
NCBI	National Center for Biotechnology Information
ORF	Open reading frame (ORF)
PCR	Polymerase chain reaction
RACE	Rapid Amplification of cDNA Ends
Rb	Retinoblastoma protein
RFC	Reduced folate carrier
RFC-1	reduced folate carrier gene
RXR	Arginine-X-Arginine motif
SAM	S-adenosylmethionine
SHMT	Serine hydroxymethyltransferase
SNP	Single nucleotide polymorphism
ssDNA	Single stranded DNA
T or t	Thymine
TBE	Tris-borate EDTA
TEA	Tetraethylammonium
TEAA	Tetraethylammonium acetate
TEAC	Tetraethylammonium chloride
THF	Tetrahydrofolate
ThTr1	Thiamine transporter
T_m	Melting temperature
TMA	Tetramethylammonium
TMAC	Tetramethylammonium chloride
TMD	Transmembrane domain
TS	Thymidylate synthase
U or u	Uracil
UDG	Uracil-DNA glycosylase

UPD	Uniparental disomy
UTR	Untranslated region
UV	Ultra-violet
V_{max}	Maximum reaction velocity
VNTR	Variable nucleotide tandem repeat

DEDICATION

I would like dedicate this thesis to those in my life I love the most. First of all to my wife, Karen, who has supported me over the years to achieve my ambitions and goals. To my two daughters, Sarah and Katy, who for the last seven years have had to put up with me saying "not this week because I am working on my PhD". My mother, who has supported everything I have wanted to do, and my father, who I still miss.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the help and support of some key people. I would like to thank Dr Julie Irving, who remembered that I wanted to study for a PhD, who supported my application and has had the patience to support me as my main supervisor. Thanks to Professor Andrew Hall, who allowed me to work as part of the prestigious Paediatric Oncology Group at the Northern Institute of Cancer Research supported by the Leukaemia Research Fund. For teaching me the techniques, providing the time to answer my questions and providing a supportive environment in which to carry out research, I would like to thank Marian Case, Lynne Minto, Elizabeth Matheson and the rest of the team in the Paul O'Gorman Building, Newcastle University.

At Northumbria University I would like to thank Dr David Holmes who at my first appraisal supported my application for Doctoral studies and set the ball rolling. Throughout the last seven years I would like to thank Professor Gary Black, Dr Mark Daniels and Dr Lisa Lee-Jones who have all taken on the role as my supervisor.

DECLARATION

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. The work was done in collaboration with the Northern Institute of Cancer Research, Newcastle University Medical School.

Name: Geoffrey Bosson

Signature:

Date: 4th May 2008

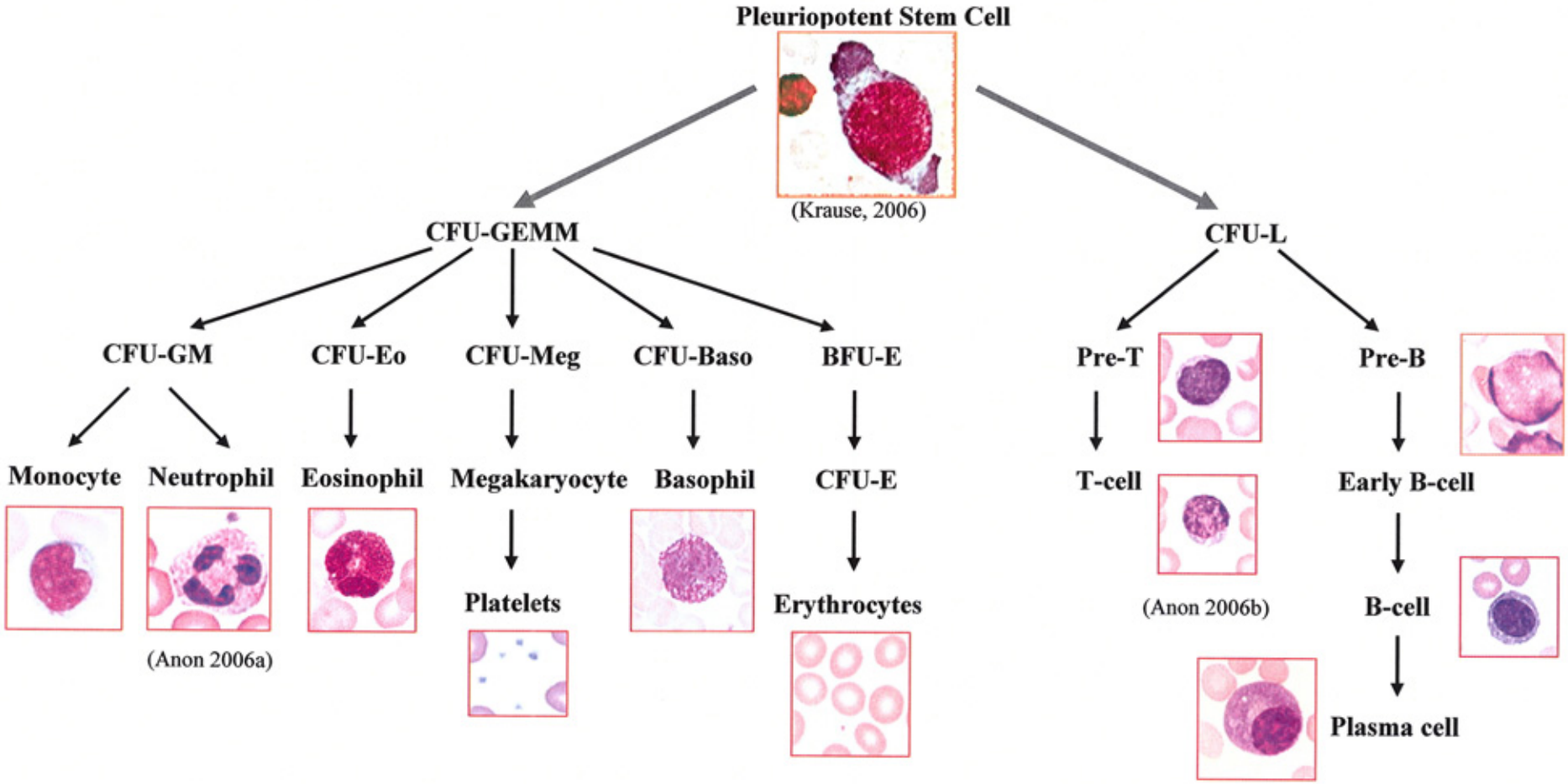
1 INTRODUCTION

1.1 ACUTE LYMPHOBLASTIC LEUKAEMIA (ALL)

Leukaemia is the result of uncontrolled proliferation of haematopoietic cells. This malignant transformation starts in a single cell and, as apoptosis and/or differentiation is lost, it replicates to produce a clone of cells that are arrested at a particular stage of development. In acute leukaemia these cells quickly replace normal bone marrow and the patient presents with anaemia, bleeding, infection and a general feeling of being unwell. Without treatment death is inevitable. The loss of differentiation can occur at any point along the haematopoietic pathway (figure 1.1) and produce either lymphoid or myeloid leukaemia.

Cells that take the lymphoid route of differentiation are active in the immune response as either T- or B-lymphocytes (McKenzie, 1996; Carotta and Nutt, 2008). T-cells released principally from the thymus of the foetus or during the first few years of life, migrate and populate lymph nodes and other lymphatic tissues. T-cells then perform one of three functions; as 'killer' or cytotoxic T-lymphocytes they bind to defective or virus infected cells and release cell membrane-destroying enzymes to initiate destruction of these unwanted cells. As T-helper cells, they assist and direct both T-cytotoxic and B-cells via the release of cytokines. To counter balance this destructive process and prevent loss of control are a third type of T-cell, the T-suppressor lymphocyte. Alternatively lymphopoesis can direct precursor cells to become B-cells which microscopically look the same as T-cells, but are ultimately responsible for production of antibodies in response to activation by specific antigens (Carotta and Nutt, 2008).

Figure 1. Haematopoeisis. (CFU, Colony Forming Unit; CFU-GEMM, colony forming unit - granulocyte, erythrocyte, macrophage, megakaryocyte; CFU-L, lymphoid; CFU-GM, granulocyte, macrophage: CFU-Eo, eosiniphil; CFU-Meg, megakaryocyte; CFU-Med, megakaryocyte; CFU-Baso, basophil; BFU-E, burst forming unit – erythroid; CFU-E, erythroid)
Unless specified, images were provided by Sysmex UK Ltd, Milton Keynes. (Geimsa stain x1000 magnification)



Few causative risk factors have been identified for the genesis of acute lymphoblastic leukaemia (ALL) (Smith *et al.*, 1999; Xie *et al.*, 2003; Pui *et al.*, 2008) but like most cancers it is likely to be the result of a complex accumulation of defects and changes. There is evidence to suggest there may be a neonatal initiation in most childhood leukaemias (Mori *et al.*, 2002; Taub *et al.*, 2002; Greaves, 2005).

Acute lymphoblastic leukaemia accounts for 23% of cancers in the under 15 age group (National Cancer Institute, 2005), making it the most common childhood cancer with 450 new cases reported each year in the UK. (Leukaemia Research Fund, 2008). It may be the most common cancer in childhood, but thankfully it responds well to modern chemotherapy and can achieve a 95% remission and 80% cure rate (Poplack, 1993; Smith *et al.*, 1999; Pui *et al.*, 2004a). However, there is still an upward trend in the incidence of leukaemia (Cancer Research UK, 2005; Xie *et al.*, 2003) and therefore a need to understand why some children fail to respond to initial treatment, or relapse, whilst others survive and live fulfilling lives (Pui *et al.*, 2008).

1.2 CLASSIFICATION

For selection of the most appropriate and effective treatment it is essential to correctly classify the cell lineage (Smith *et al.*, 1996). Pui and Evans (1998) provide an algorithm for classification into high, standard and low risk group, but as our understanding of leukaemia increases more factors can be integrated into this risk stratification. Classification used in the Medical Research Council trials (Hann *et al.*, 2001) and by the Children's Oncology Group (Carroll *et al.*, 2003) is based on:-

1.2.1 Clinical features

1.2.1.1 **Age at presentation** – Presentation in the first year of life or in adolescence has a less favourable outcome than those who present between ages 1-9 years (Smith *et al.*, 1999; National Cancer Institute, 2005; Pui *et al.*, 2008).

1.2.1.2 **White blood count at diagnosis** – A cut-off value of 50,000/μL of blood is usually used to classify patients, with counts greater than this value associated with poor prognosis. However a single cut-off value does not convey that this is a continuous scale with significantly higher counts having worse prognosis (Sorich *et al*, 2008).

1.2.1.3 **Cerebrospinal fluid white cell count** – A white cell count ≤5/μL of cerebro-spinal fluid (CSF) indicates infiltration to potential sanctuary sites and therefore requires more intensive therapy. (Chessells *et al.*, 1995).

1.2.1.4 **Gender** – Cases of relapse due to leukaemic cell sanctuary in the testes have almost been eliminated by modern therapy and yet the prognosis is still poorer in boys (Chessells *et al.*, 1995; Pui *et al.*, 1999). Pui *et al* (1999) indicated that whilst the cause is likely to be multifactorial, the incidence of T-cell lineage is higher in boys and being inherently harder to treat may account for some of this poor prognosis. Therapy continues to improve (Pui *et al.,* 2004b) and poor prognosis in males is less significant today (Pui *et al*, 2008).

1.2.1.4 **Race** – The SEER report (Smith *et al.*, 1999) indicates that the initial incidence is lower in black children but the chance of relapse is higher. Kadan-Lottick *et al.,* (2003) caution that racial and ethnic differences could be the result of willingness of particular groups to participate in studies.

1.2.2 Characteristics of the leukaemic cells

The cell type and its degree of maturity also affect the prognosis, with T-cell ALL having poorer prognosis than B-cells (Kager *et al.*, 2005) so classification of the cell line is required based on the following:-

1.2.2.1 **Morphology** – Historically the French-American-British (FAB) classification system identified three types of ALL based on the microscopic assessment of cellular maturity (Bennet *et al.*, 1981). Due to its subjectivity, the FAB classification has been dropped in the UK with greater reliance being placed on the following criteria.

1.2.2.2 **Immunophenotype** – This uses the presence of cytoplasmic or cell surface markers, the latter usually referred to as 'clusters of differentiation' (CD) that can be detected immunologically. As the lymphocytes mature, the pattern of expressed cytoplasmic and CD markers change allowing classification of lymphocytes rapidly by flow cytometry into early pro-B, pre-B, mature B and T-cells.

1.2.2.3 **Cytogenetics** – Cytogenetic changes are seen in 80% of children with ALL (Faderl *et al.*, 2003) and can be classified as either a gross change in the number of chromosomes (ploidy); and/or movement of genetic material between chromosomes (translocation) causing abnormal transcription of critical genes. Hypodiploidy (<45 chromosomes) is associated with adverse risk and poor outcome (Heerema *et al.*, 1999), but thankfully only occurs in 1% of childhood ALL cases (Hann *et al.*, 2001). Conversely, hyperdiploidy (>50 chromosomes) is associated with superior prognosis (Hann *et al.*, 2001, Moorman *et al.*, 2003; Kager *et al.*, 2005; Pui *et al.*, 2008), especially in children with acute myeloid leukaemia (AML). This is thought to be the result of a higher expression of key enzymes and other proteins involved in folate metabolism, and goes some way to explain the higher survival and lower relapse rate

of patients with hyperploidy. In ALL, the chromosome most frequently seen in increased numbers is chromosome 21 (Raimondi *et al.*, 1992; Moorman *et al.*, 2008) and although it shows a different pattern of gene expression than in AML it provides a similar improved prognosis (Raimondi *et al.*, 1992; Whitehead *et al.*, 1992; Ito *et al.*, 1999; Belkov *et al.*, 1999, Pui *et al.*, 2008). As chromosome 21 appears to play such a significant and beneficial role it would be expected that Down's syndrome (trisomy 21) would be considered a favourable prognostic characteristic. In AML it is, but in ALL, outcome is poorer and the chance of relapse higher (Chessels *et al.*, 2001; Taub and Ge, 2005). For this reason Down's Syndrome should be considered as a risk factor in determining the best course of treatment (Zipursky *et al.*, 1997).

Four major translocations have been identified that have significant influence on the effectiveness of treatment.

- t(1;19)(q23;q13) is mainly found in pre-B ALL and results in the E2A and PBX gene fusion (Kager *et al.*, 2005). Prognosis in this group is improving with changes in chemotherapy (Carroll *et al.*, 2003).
- Mixed Lineage Leukemia (MLL). Translocation or deletion of the q23 band on chromosome 11 onto a number of other chromosomes occurs in 8% of ALL cases (Pui *et al.*, 2002). This heterogenous group of abnormalities is found in both myeloid and lymphoid lineage and tend to have poor clinical outcome, especially in children less than one year old (Pui, 2005; Harrison *et al.*, 2007).
- t(9;22)(q34;q11) creates the *BCR-ABL* gene, also known as the Philadelphia chromosome. Whilst only found in 3-5% of cases of childhood ALL, the increased tyrosine kinase activity generates cell lineage which is generally more difficult to treat. (Arico *et al.*, 2000; Chiaretti *et al.*, 2007)

- t(12;21)(p13;q22) is the most common translocation in childhood ALL, present in 20-25% of B-cell ALL, (Asakura *et al.*, 2004; Tsuzuki *et al.*, 2007). This brings together elements of the TEL gene on chromosome 12 and the AML-1 gene on chromosome 21 creating the *TEL-AML1* gene and facilitates expression of the *TEL-AML1* protein. Whilst the *TEL-AML1* gene is thought to be the result of a prenatal initiation event that increases the risk of ALL by 100 fold (McHale *et al.*, 2003), it is still considered a favourable prognostic marker for successful treatment (Seeger *et al.*, 1999; Morrow *et al.*, 2004; Zuna *et al.*, 2004; Tsuzuki *et al.*, 2007; Davies *et al.*, 2008). Whitehead *et al* (2001) show that there are higher concentrations of polyglutamylated MTX in cells with this translocation and Asakura *et al* (2004) show that one reason could be that the *TEL-AML1* protein represses transcription of the multidrug resistance (MDR) gene reducing the ability of the cell to remove MTX.

1.2.2.4 **Gene expression** – The biggest breakthrough for rapid classification is the microarray chip ('gene chip') that can screen for expressed genes (Gottardo *et al.*, 2007; Lu *et al.*, 2007). This can discriminate between AML and ALL, T-cell and B-cell and subtypes (Mi *et al.*, 2007; Kuchinskaya *et al.*, 2008). Their potential lies in the possibility that they can be used in the clinic situation (point-of care testing) for immediate risk management (Harrison *et al.*, 2007).

1.2.3 Early response to treatment – As would be expected, those patients who respond well to induction therapy have a better outcome. The success of early treatment is often assessed by sensitive molecular techniques that can detect residual disease down to 10^{-4} (Donadieu & Hill 2001; Hoelzer *et al.*, 2002; Digiueseppe 2007; Jolkowska *et al.*, 2007).

1.3 TREATMENT

Over the years, clinical trials for the treatment of ALL have built on the information of earlier trials resulting in therapy regimen that provide the success and hope that can be offered today (Chessells *et al.,* 1995; Hann *et al.,* 2001; MRC UKALL 2003; Pui *et al.*, 2008). Whilst each trial may introduce a new drug or modify when the multi-drug therapy is given, there is usually a 3 stage approach:-

1.3.1 **Induction of remission** – course of drugs that include allopurinol, daunorubicin, dexamethasone, vincristine, asparaginase, 6-mercaptopurine (6-MP) started at diagnosis and given over a four week period, plus intrathecal methotrexate and cytarabine. If successful, the full blood count should return to normal within 4 weeks, but as mentioned in section 1.2.3, earlier remission assessed by minimal residual disease is indicative of good prognosis.

1.3.2 **Intensification (or consolidation)** – once a normal blood profile has been regained a regimen of chemotherapy cycles, designed to ensure any residual disease is eliminated, is started and continued for several weeks at a time. Periods of rest are usually required between each cycle of drugs, which in the current MRC ALL trial 2003 (MRC UKALL, 2003) consists of 6-MP, cyclophosphamide, cytarabine, vincristine, asparaginase and further intra-thecal methotrexate. The exact combination depends on which regimen the patient has been allocated.

1.3.3 **Continuation (or maintenance) therapy** – during this period a combination of drugs at lower doses (usually given orally) are used over 2-3 years (2 for girls, 3 for boys) to ensure the malignant lineage does not return. Vincristine, 6-MP, dexamethasone, oral and intrathecal methotrexate are used in the 2003 trial.

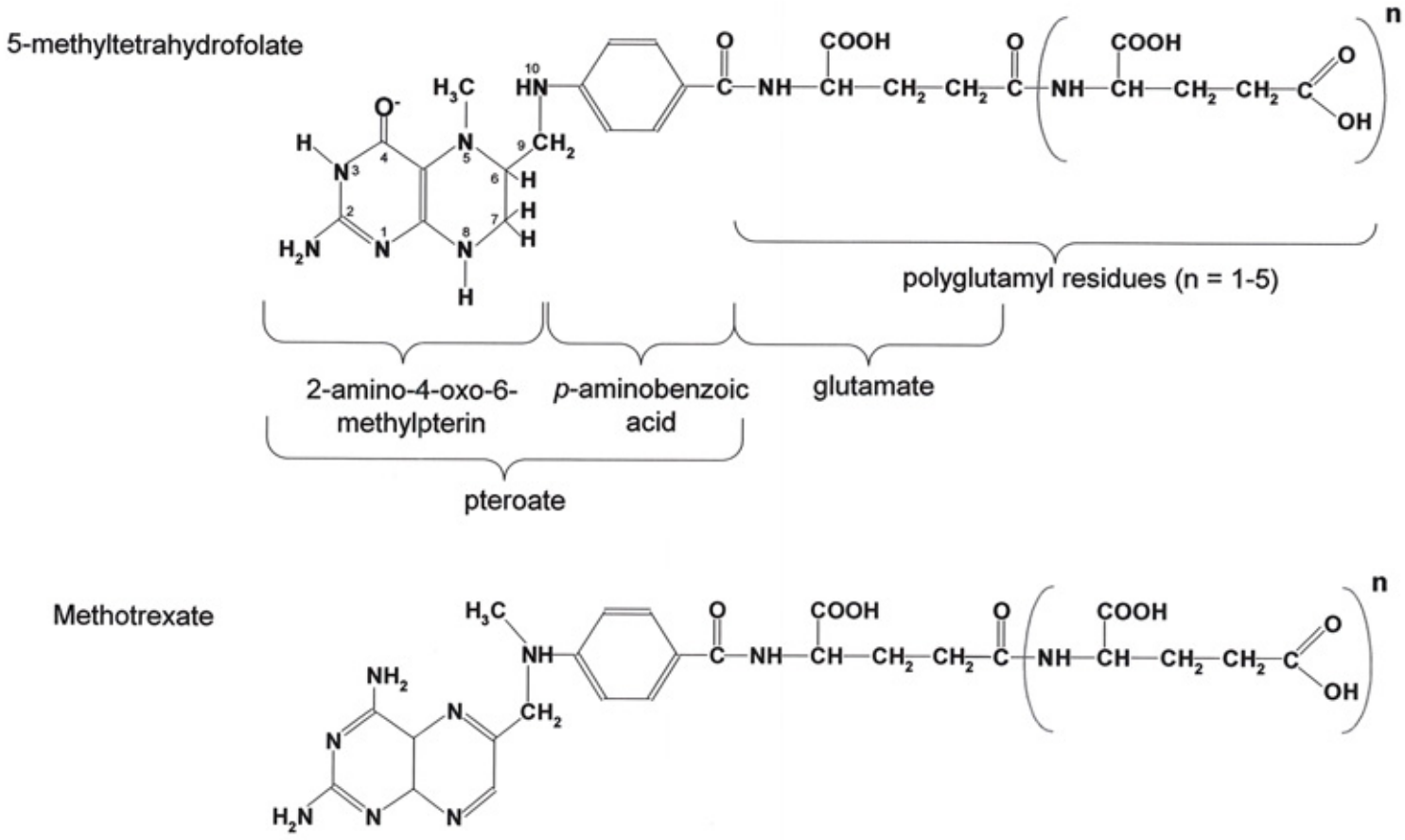
Ideally these drugs will selectively cause apoptosis of the malignant cells and suppress any further growth without having any effect on normal cells. Unfortunately this is not the case and side effects are inevitable when using such toxic compounds. These side effects, including nausea, convulsions and other neurotoxic effects (Kishi *et al.*, 2003; Mittal *et al.*, 2005), diarrhoea, tiredness, bleeding, hair loss, poor healing and growth abnormalities can often be as distressing to the patient and their family as the leukaemia itself. Even if successful at treating the primary cancer, the drugs may initiate a secondary cancer in up to 3% of cases later in life (Gadner *et al.*, 2006; Hijiya *et al., 2007;* Ebeo *et al.*, 2008). The challenge for the clinician is to determine the most effective dose to achieve a cure and yet keep the side effects to a minimum. As will be seen in the coming pages there is a 'trade-off' as inadequate chemotherapy can result in drug resistance and relapse. Such resistance is more likely if malignant cells have infiltrated sanctuary sites, such as the central nervous system and the testes, where standard drug therapy may not produce sufficiently high concentrations to be pharmacologically active. For this reason, intrathecal administration of methotrexate is routinely used.

1.3.4 Methotrexate – Following World War II, it was noted by clinicians that the cell count in leukaemia patients improved when the diet was deficient in folic acid. Sidney Farber believed that if a drug could be found to interfere with folate metabolism then this effect could be enhanced and form the basis of treatment (Farber *et al.,* 1948). Aminopterin, shown to inhibit this key metabolic pathway, had just been discovered and was undergoing trials. Farber reported his results in 1948, but because of some severe side-effects he started the search for a less toxic replacement and selected another folate antagonist, methotrexate; *N*-[4-[[(2,4-diamino-6-pteridinyl)-methyl]-methyl-amino]benzoyl]-L-glutamic acid.

Since its discovery, methotrexate has become a mainstay of chemotherapy in many cancers and at lower doses to treat inflammatory disease (Jolivet *et al.*, 1983, Rots *et al.,* 2000a; Urano *et al.*, 2003; de Lathouder *et al.*, 2003). Whilst inhibition of folate metabolism affects all cells, it is those that have a high replication rate (eg., malignant cells) and thus greatest demand for reduced folates which are most susceptible. This also explains many of the side effects which are the result of impaired replication of tissues with a normally high rate of cell turnover (eg., intestinal epithelium) (Belur *et al.*, 2001) or due to disturbances in homocysteine remethylation to methionine (Sigmund *et al.*, 2003 Kishi *et al.*, 2003). To give normal cells a chance to recover during high-dose MTX therapy, leucoverin (N^5-formyl-tetrahydrofolate) can be given at specific times during treatment to replenish the intracellular tetrahydrofolate pool (Asai *et al.*, 2003). This works because normal cells have the enzyme methionine synthase to convert N^5-formyltetrahydrofolate to N^5N^{10}-methylenetetrahydrofolate, thus by-passing the need for dihydrofolate reductase (DHFR) to generate tetrahydrofolate (THF). This allows higher, and therefore more effective doses of MTX to be used (Hamel *et al.*, 1981; Synold *et al.*, 1994).

The pharmacological effect of methotrexate is due to its structural similarity to the natural folate (figure 1.2) and its action as a competitive inhibitor at several key points along the folate biochemical pathway.

Figure 1.2. Structures of methotrexate and tetrahydrofolate showing the similarity that allows methotrexate to act as a competitive inhibitor. (Modified from Jolivet *et al.*,1983 and Campbell *et al.*, 2005)



1.4 FOLATE METABOLISM

As far as we know, all life on Earth is dependent on the carbon-based biomolecules; carbohydrates, amino acids, lipids, nucleic acids, etc. Such organic molecules can be metabolised to form a potentially infinite number of compounds many of which are essential for the cell's well being. The *de novo* synthesis of biomolecules is principally reliant on the sequential addition of single carbon molecules and therefore on a series of one-carbon (1-C) donors. The most important of these donor molecules is folate which is found in all cellular life-forms and plays a key role in the *de novo* synthesis of nucleic acids, making them essential for cell replication. However, not all cells have the ability to synthesise folate and must obtain folates from outside the cell. Human cells fall into this group and we must obtain folates from the diet, hence the classification of folate as a vitamin. A normal balanced diet provides sufficient folate, but the richest sources are green-leaf vegetables.

Folates are a group of related molecules based on, and including, folic acid. Folic acid is the oxidised form that is reduced to dihydrofolate (DHF) and then further reduced to the fully active tetrahydrofolate (THF) by dihydrofolate reductase (figure 1.3). This then allows it to acquire and hold the 1-C residue which undergoes conversion into a range of molecules (figure 1.4) that now act as co-substrate in several important reactions (figure 1.5). The majority of dietary folate exists in a polyglutamylated form and first requires removal of glutamyl residues by glutamate carboxypeptidase II at the brush border before it can be absorbed (Morin *et al.*, 2003).

Figure 1.3. Biochemical pathway showing activation of oxidised folic acid to fully reduced tetrahydrofolate (THF) by dihydrofolate reductase.
(Modified from Horton *et al.*, 2006).

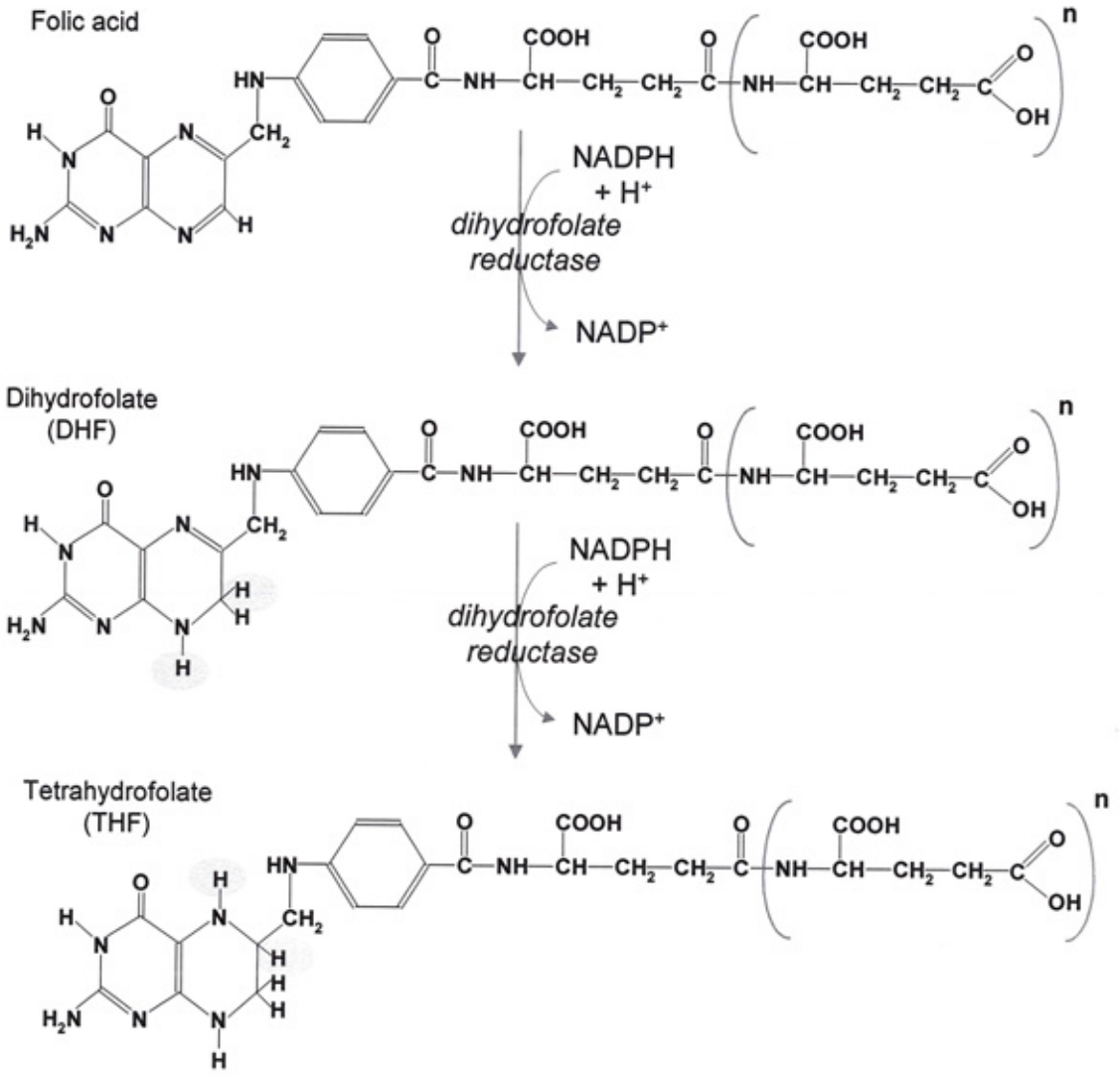


Figure 1.4. Biochemical pathway showing conversion of fully reduced tetrahydrofolate (THF) to the various 1-carbon donor states. For clarity, only the pterin residue is shown in each intermediary, the remaining PABA-glutamate residue seen in figures 2 and 3 are represented by R. (Modified from Voet *et al.*, 2006).

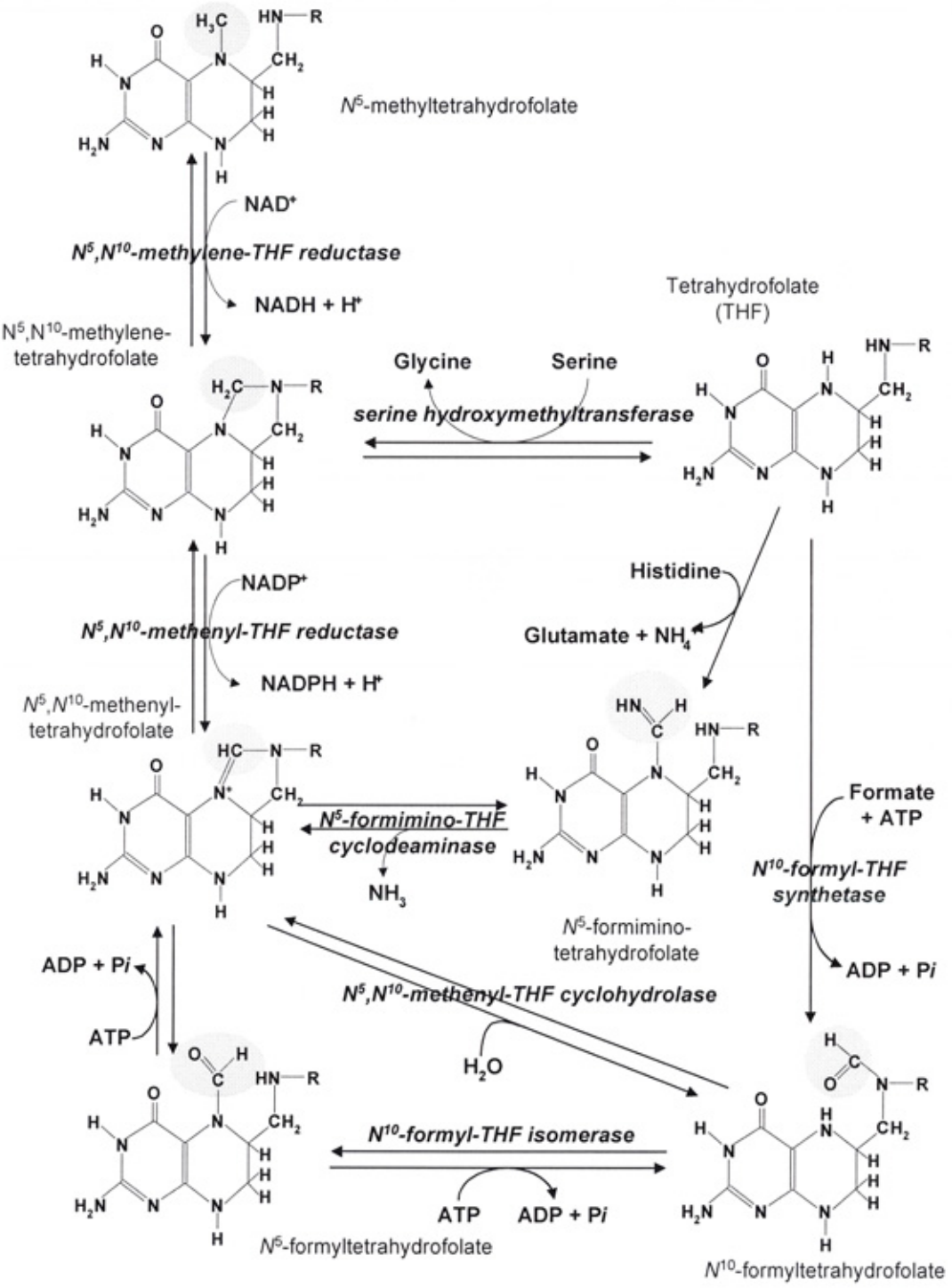
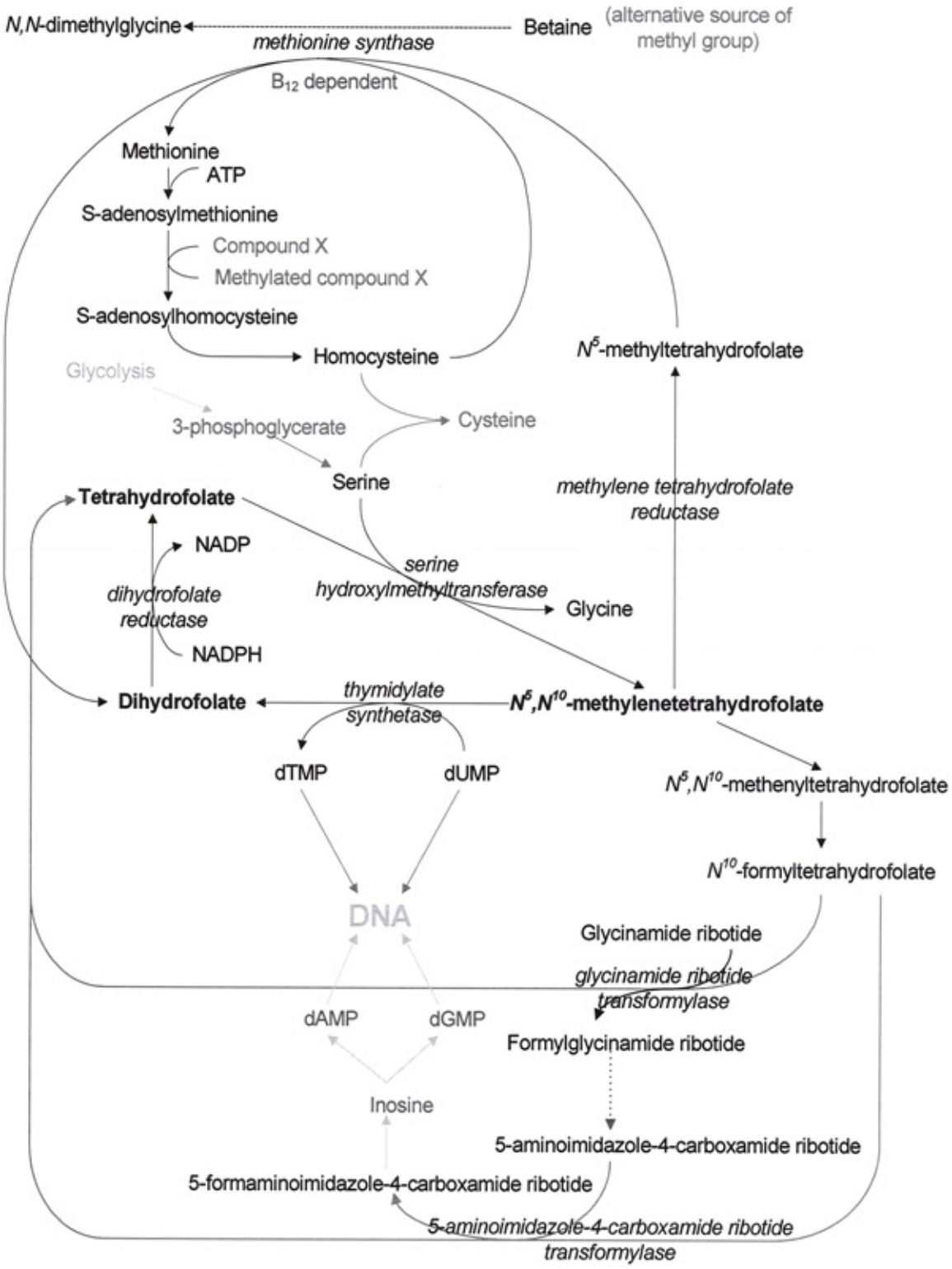


Figure 1.5. Diagram showing inter-related pathways requiring 1-carbon donation from folates. Enzymes are shown in italics. (modified from Jolivet *et al.,* 1983; Salooja *et al.*, 1998; Selhub, 1999; Carroll *et al.*, 2003; Nijhout *et al.*, 2004, Kohlmeier *et al.*, 2005).



The folates are then absorbed along the small intestine using the intestinal folate carrier (IFC-1) which transports them across the brush border to the basolateral membrane. Here they are thought to diffuse along a concentration gradient into the submucosa (Wang *et al.*, 2000), but there is increasing evidence from animal studies (Zeng *et al.,* 2000) that an export mechanism at the basolateral membrane actively pumps folates from the cell and enhances folate uptake from the intestinal lumen to replace that moving out of the cell. In either case, the folates delivered into the submucosa will pass into the portal circulation and thence to the rest of the body. The higher demand for folate of the growing foetus is met by transfer from the maternal circulation via a placental folate transporter system (Bisseling *et al.,* 2004).

The major circulating folates are the N^5-methyltetrahydrofolate (5-methylTHF) and N^5-formyltetrahydrofolate (5-formylTHF) forms (structures shown in figure 1.4). At high extracellular concentrations these can simply diffuse into the cells, but because folates are essentially hydrophilic molecules and do not readily diffuse across cell membranes this does not provide sufficient intracellular folate for normal cell metabolism. To provide an adequate intracellular folate concentration, active transport mechanisms must be used that can work against a concentration gradient.

1.4.1 Folate receptors. Folate receptors (also known as folate binding proteins, FBP) are glycosylphosphatidylinositol membrane bound proteins with a high affinity for folic acid (K_M 1-10 nM) (Drori *et al.,* 2000a; Mauritz *et al.*, 2008) and 5-methylTHF (K_M 1-5 nM) (Trippett *et al.,* 1992). When folic acid or 5-methylTHF binds to a receptor the whole complex is internalised against the concentration gradient by endocytosis (Wang *et al.,* 1992; Brigle *et al.,* 1994; Trippett *et al.,* 1999; Khokhar, 2002). Although this process is slower and requires more energy than the

RFC route, the receptors are thought to play a significant role in renal reabsorption (Weitman *et al.,* 1992; Matherly and Goldman, 2003).

1.4.2 Reduced folate carrier (RFC). The folate receptors described above have a low affinity for reduced folates (K_M 10-300nM) (Trippett *et al.,* 1999) and these more abundantly available folates must therefore be brought into the cell using a different mechanism, the reduced folate carrier (RFC). This bi-directional anionic exchanger is the most important of the folate uptake mechanisms with high affinity for reduced folates (Drori *et al.,* 2000a) and will be discussed in detail later (section 1.7).

The two systems may be coordinated to provide sufficient intracellular folates (McKusick, 1995), whilst others believe they act independently (Westerhof *et al.,* 1991; Dixon *et al.,* 1992). There also appears to be a third mechanism which is able to recognise all folates with almost equal affinity (K_M 1-5μM) and operates at low pH (Sierra and Goldman, 1999; Zhao *et al.*, 2005).

1.4.3 Low pH dependent transporter. Although its role is still poorly understood, a third mechanism has been proposed (Drori *et al.,* 2000a; Rajgopal *et al.,* 2001) with a high efficiency for folate transport at the low pH achieved by the micro-environment at the luminal surface of the intestinal cells. Rajgaopal *et al.,* (2001) suggested that this was a post-translation splice variant of RFC, but has since been proven to be independent of the reduced folate carrier (Wang *et al.*, 2004; Wang *et al.*, 2005) and awaits further investigation and understanding.

Once inside the cell folylpolyglutamate synthetase (FPGS) adds glutamate residues to the C-terminal of the existing single glutamyl residue (Liani *et al.*, 2003). This is to increase its size and polarity to prevent its loss from the cell via the folate export

pumps (Sierra and Goldman, 1999). Folic acid must then be reduced by the enzyme dihydrofolate reductase (DHFR) in a 2-step process (figure 1.3) to produce the fully active tetra-hydrofolate (Wang & Fenech, 2003). Note that the same enzyme catalyses both steps, but because the preferred substrate is dihydrofolate, then the first step is slower than the second (Horton *et al.,* 1993). Thus, this enzyme plays a pivotal role in the activation and regeneration of folates and will be discussed in more detail later (section 1.10).

Folates can be excreted from the cell via the multi-drug resistant protein family (MRP) but first require removal of the polyglutamyl residues by folylpolyglutamyl hydrolase (γ-glutamyl hydrolase) (Galivan *et al.,* 2000). The MRP exists in five known isoforms, MRP 1–5 (Chen *et al.*, 2002), which all use ATP to facilitate movement of folates and antifolates out of the cell. Although the carrying capacity of the MRP and affinity for folates and antifolates is low, the high expression of MRP seen in some circumstances can expel them from the cell before they have a chance to become polyglutamylated and in the case of antifolates produce a degree of resistance (Hooijberg *et al.,* 1999; Zeng H, 2001; Kruh *et al.*, 2001; Chen *et al.*, 2002). It is believed the MRP may play a role in the control of intracellular folate concentration (Hooijberg *et al.*, 2003), ensuring that the enzymes of folate metabolism are operating most efficiently and concentration of free intracellular folate is kept low

As described earlier, the fully reduced THF now accepts a 1-carbon group which it can hold in a variety of oxidative states, methyl ($-CH_3$), methylene ($-CH_2$), methenyl (-CH=), formyl (-CH=O) or formimino (-CH=NH) groups (figure 1.4). Under thymidylate synthetase catalysis, one of the THF intermediates, N^5,N^{10}methyleneTHF, donates its methylene group to deoxyuridine monophosphate (dUMP) and generates

the cell's source of *de-novo* deoxythymidylate monophosphate (dTMP). The remaining 7,8-dihydrofolate is recycled back to THF using DHFR (figure 1.5).

The mitochondria also have a requirement for folates and obtain reduced folates from the cytoplasmic pool using an independent reduced folate carrier in the mitochondrial wall (Trippett *et al.,* 2001).

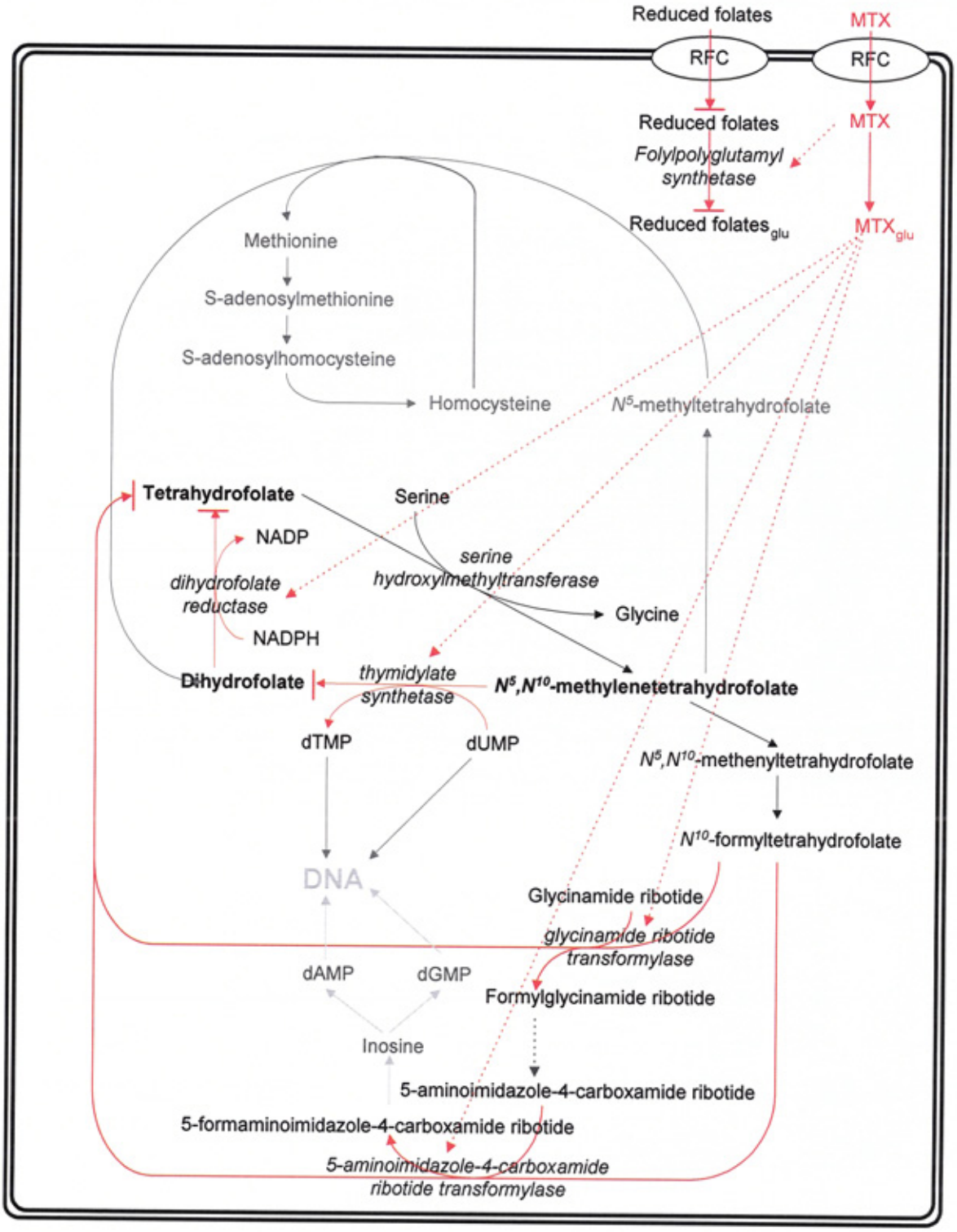
The importance of folates should now be evident and is further emphasised by reports of deficiency associated disease such as megaloblastic anaemia, neural tube defects and recently cardiovascular disease (Salooja *et al.*, 1998; Chango *et al.*, 2000; Toole *et al.*, 2004).

1.5 ANTIFOLATE EFFECT ON FOLATE METABOLISM

The competitive properties of MTX are evident at several points in folate metabolism (figure 1.6). Whilst MTX can diffuse across the cell membrane at very high concentrations it tends to use the reduced folate carrier as its main route of entry into the cell. The reduced folate carrier is a rate limiting process and in the presence of extracellular MTX the normal uptake of reduced folate is proportionally reduced.

Once inside the cell MTX first competes with intracellular folate for FPGS with two effects. First, the true folates are no longer polyglutamylated and this makes their excretion from the cell more likely. Secondly, MTX is polyglutamylated in its place making it more difficult for the cell to excrete it from the cell and increases the potency of the drug (Jolivet *et al.*, 1983). As the intracellular ratio of MTX to natural folates increases, the competitive nature of MTX is more effective.

Figure 1.6. Diagram showing uptake of methotrexate and inhibition of key enzymes in folate metabolism. Cell membrane indicated by double thick line. Points of competitive inhibition by MTX and polyglutamylated MTX (MTX_{glu}) shown in red. (Adapted from Carroll *et al.*, 2003).



Inhibition of DHFR quickly depletes the intracellular pool of THF as DHF can no longer be recycled in sufficient quantities. This alone would have an effect on cell metabolism, but the fact that polyglutamylated MTX is also inhibiting TS (Chu *et al.*, 1990), glycinamide ribotide transformylase (GART) and 5-aminoimidazole-4-carboxamide ribotide transformylase (AICART) makes it even more effective. As GART and AICART are inhibited, the *de novo* synthesis of purines (adenine and guanine) is impaired leaving the cell dependent on its ability to recycle purines. Inhibition of TS reduces the ability of the cell to synthesise *de novo* thymidine (Chu *et al.*, 1990; Hornung *et al.*, 2000). The overall effect is a reduction in the ability for DNA replication and cell proliferation (Belur *et al.*, 2001).

Under normal circumstances the cell maintains a relatively low level of dUTP by using deoxyuridine triphosphate nucleotidohydrolase (dUTPase; EC 3.6.1.23) to reduce the chance of its misincorporation during DNA replication or repair (Goulian *et al.*, 1980). The enzyme catalyses the release of pyrophosphate from dUTP producing the monophosphate form which no longer has the capacity to act as the substrate for incorporation. If uracil is inadvertently used and incorporated, then uracil-DNA glycosylase (UDG; EC 3.2.2.3) catalyses the excision of the offending uracil and allows it to be replaced by the correct thymine. However, if the cell is deficient in dTTP the nick in the DNA is still made and the offending base removed, but cannot be repaired. When this occurs on both strands of the double helix, one in either strand within 14 bases of each other, then a double-strand break is likely to result (Blount *et al.*, 1997) with consequent break-up of the DNA, formation of micronuclei (Jacky *et al.*, 1983; Wickramsinghe and Frida, 1994; MacGregor *et al.*, 1997), production of missense codons (Rots *et al.*, 2000b) and what is frequently referred to as 'thymineless death'.

This scenario is induced by MTX (Rots *et al.*, 2000b). Mature cells have little demand for THF and are minimally affected by MTX, but those cells undergoing rapid replication (eg., cells of the normal bone marrow, gastrointestinal mucosa, as well as cancer cells) will take up the drug when their demand is highest and effectively poison themselves (Hoffbrand *et al.,* 1973).

1.6 METHOTREXATE RESISTANCE

Natural selection in living organisms is based on the fact that mutations in a cell's genome may give it a survival advantage in a modified environment. This means that if cells are exposed to MTX levels insufficient to kill them outright, they will have a chance to mutate. If the mutation renders the cell less sensitive, or even insensitive to MTX, then the patient relapses and current treatment is no longer effective (Jolivet *et al.,* 1983; Ohnuma *et al.,* 1985; Guo *et al.*, 1999; Rots *et al.*, 2000a). Alternative regimens using different drugs or higher doses of current drugs will be required, but in either case relapse survival outcome is still poor. For this reason, it is important to identify and understand in each patient the mechanism or mechanisms of resistance that caused relapse, so that the most appropriate therapy can be started and prognosis for these 20-30% of cases improved (Heerema *et al.*, 1999; Cheok and Evans, 2006).

Relapse due to MTX resistance can result from mutations affecting cellular uptake, cellular efflux, intracellular activation to the fully active form, affinity of target enzymes, or a modified apoptopic threshold. As MTX will only be able to inhibit target enzymes once inside the cell, then the first of these mechanisms, (ie., loss of RFC activity) may be considered the most effective (Gong *et al.,* 1997; Trippett & Bertino, 1999; Sadlish *et al.,* 2000). Simple down regulation of the protein or loss of functionality would be effective, but this would also reduce the uptake of reduced

folates still required by the malignant cell to survive and replicate. Therefore, a more subtle change in the carrier protein to selectively transport reduced folates rather than MTX would be even more effective (Rots *et al.,* 2000b). Changes to the reduced folate carrier observed in a variety of cell lines (*in vitro*) will be discussed fully in section 1.7. If MTX does gain entry into the cell, then the second line of defence would be to expel it as soon as possible and keep the intracellular concentration so low it is ineffective as a competitive inhibitor. The ubiquitous multidrug resistance proteins (MRP) are responsible for expulsion of a variety of xenobiotics, including MTX, and to do this efficiently requires the antifolate in its non-polyglutamylated form. Therefore, preventing FPGS catalysed polyglutamylation of MTX (Wang *et al.*, 2003) and/or increasing depolyglutamylation by FPGH (Rots *et al.,* 2000b) are mechanisms which have been noted *in vitro* (Rots *et al.*, 1999) and will be further discussed in sections 1.8 and 1.9 respectively. If none of these mechanisms are evident, then resistance can still be achieved at the target enzyme itself by simply increasing the amount of the target enzyme (Ohnuma *et al.,* 1985; Rots *et al.*, 2000c) and/or modifying the selective nature of the enzyme (Peters & Jansen, 2001). In the former, there is sufficient enzyme to bind MTX and still leave sufficient uninhibited enzyme to perform the reaction essential for the malignant cell to survive (Rots *et al.*, 2000b). The latter uses changes to the amino acid sequence in the active site to increase the affinity for true substrate or reduces affinity for MTX.

The overall action of MTX is to cause 'thymineless death' (ie., apoptosis induced by irreparable damage to the DNA). However, it is not uncommon for the 'apoptic threshold' to be reset at a higher level in malignant cells as a result of changes to the Bcl-2 family of oncoproteins. Under normal circumstances there is a balanced ratio of apoptopic suppressors, such as Bcl-2 and Mcl-1; and apoptopic promoters, such as

Bax and Bcl-xs (Findley *et al.*, 1997). In several cancers an increase in Bcl-2 has been shown to confer resistance to a whole range of anticancer drugs as apoptosis is suppressed and damage caused by the drug is tolerated (Hannun, 1997). Whilst the increased expression of the Bcl-2 protein is a well recognised mechanism for drug resistance, including resistance to MTX, it is not a common mechanism causing relapse in childhood ALL (Findley *et al.*, 1997; Hogarth and Hall, 1999; Wall *et al.*, 1999). In the investigation of apoptosis in childhood ALL by Hogarth and Hall (1999) the correlation with relapse was strongest with expression of Bax alone.

Several groups throughout the world have investigated, and continue to investigate the mechanisms leading to drug resistance and relapse, including the Paediatric Oncology Group at the Northern Institute of Cancer Research in Newcastle-upon-Tyne. Support from the Leukaemia Research Fund (LRF) is allowing this group to look at MTX and other drugs used in the treatment of ALL to understand more about the failure of therapy, development of resistance and provide information that will allow selection of the most effective treatment. The research detailed in this thesis increases the understanding of DHFR and RFC at the molecular level and adds to the overall aim of the group. The remainder of the introduction is intended to provide an overview of the biochemical importance and outline molecular biology of RFC and key enzymes in folate metabolism.

1.7 REDUCED FOLATE CARRIER (RFC)

1.7.1 PROTEIN STRUCTURE AND FUNCTION

Although much of the early research on RFC was performed on mouse and hamster models, the common need for folate uptake by mammalian cells means that evolution has produced a protein which is very similar across the species (Bosson, 2003). This has allowed some important information on RFC in humans to be gained from animal models (Zhang *et al.*, 1998a; Williams & Flintoff; 1995; Moscow *et al.,* 1995).

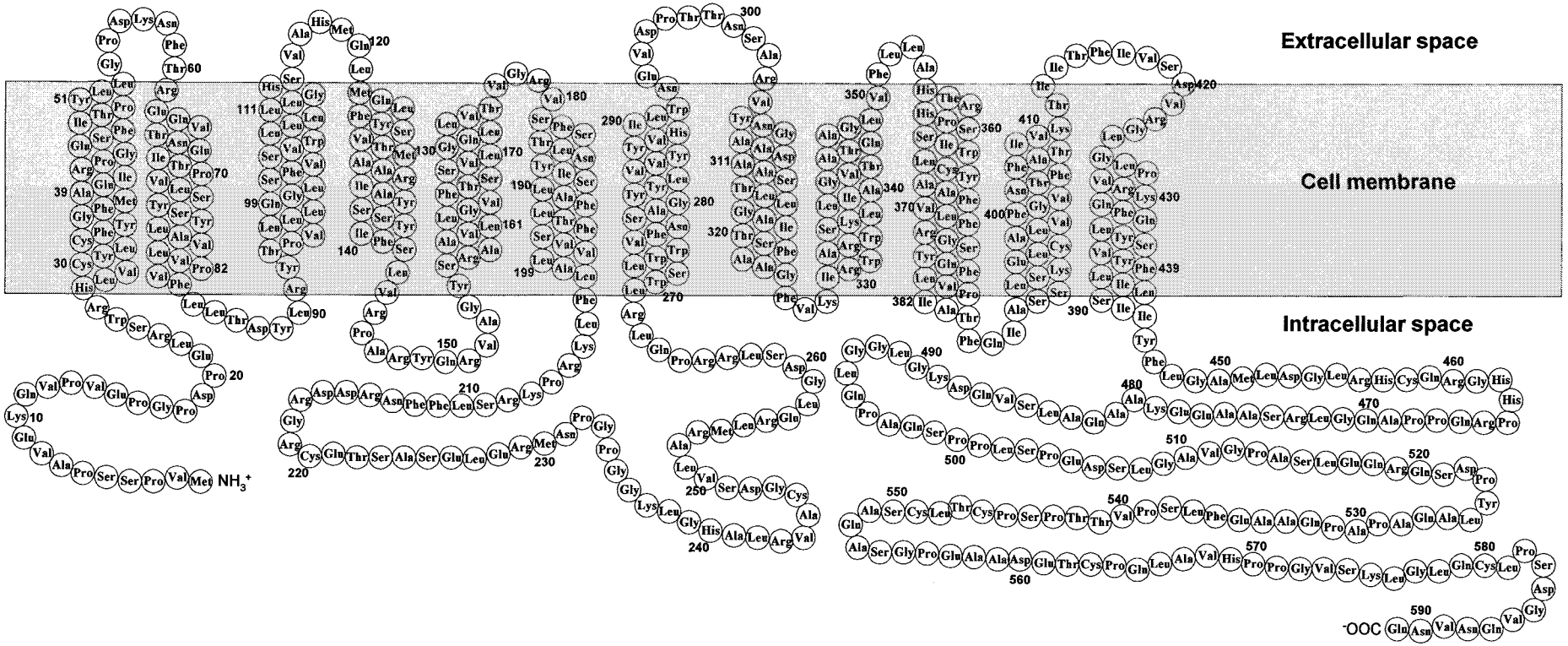
The 591 amino acid protein is highly glycosylated in humans increasing its size from 65,000 Dalton to a mass of 92,000 (Zhang *et al.*, 1998a). Functionally, RFC facilitates anion exchange bringing hydrophilic folates across the cell membrane using the negatively charged glutamate residue of folate (Pao *et al.*, 1998). Such a system requires positively charged amino acids and the highly conserved arginine residues in the protein are likely candidates (Zhao *et al.,* 1999). Whilst the charge of these amino acids is important, Zhao *et al. (1999)* showed that by substituting Arg-131 on murine RFC with histidine (replacing one positively charged amino acid with another), activity was still lost and therefore charge is not the only factor.

The protein structure has a conserved potential phosphorylation site for protein kinase-C at residue 23 in all three species (Moscow *et al.,* 1995), but its use in activating or driving transport, has not been indicated by any research group. Neither does RFC show any structural motifs typical of transmembrane transport proteins that utilise ATP. This leaves the theory that intracellular phosphate is the driving force for folate uptake (Goldman, 1971; Sierra & Goldman, 1999). The high intra-cellular phosphate (PO_4^{2-}) is the anion leaving the cell along a concentration gradient, but to maintain neutrality an anion must also move into the cell. The extra-cellular

phosphate concentration is low and is therefore unlikely to move back into the cell. RFC is selectively holding a negatively charged folate on the extracellular surface and in effect allows it across the membrane to counterbalance the charge (Goldman, 1971).

The structure of RFC shown in figure 1.7 is based on the Kyte-Doolitle hydropathy plot (Dixon *et al.,* 1994) and the Hopp and Wood hydrophilicity plot (Trippett, 2001). The putative transmembrane arrangement derived from this data is that accepted by the major RFC research groups (Drori *et al.,* 2000a; Sierra & Goldman, 1999; Zhao *et al.,* 1999; Liu & Matherly 2001; Matherly, 2001), however the complete 3-dimensional structure is still to be confirmed. Figure 1.7 clearly shows the proposed arrangement of 12 α-helical transmembrane domains made up of hydrophobic amino acids. Transmembrane domains (TMD) range in size from 17-25 residues and are highly conserved between the 3 species studied (Wong *et al.,* 1995), especially the 9 tryptophan residues and 4 of the 6 cysteine residues (Wong *et al.,* 1995). Hydrophilic amino acids make up the four small intracellular loops, six extracellular loops, C-terminal, N-terminal regions and the large intracellular loop formed half-way along the sequence. Both C and N-terminals of normal RFC are shown lying in the cytoplasm. This was confirmed experimentally by the transfection of a known epitope construct into the C-terminal of a K562 cell line and subsequent localization of the translated product using immunofluorescent staining for the inserted epitope (Ferguson & Flintoff, 1999). The fluorescent label was bound to the cell membrane, but only after the cell was made permeable with Triton X-100 (Ferguson & Flintoff, 1999), therefore the terminal ends must reside inside the cell. Similar studies have been done to confirm the position of other domains in relation to the cell membrane (Sadlish *et al.,* 2002a; Cao & Matherly, 2004).

Figure 1.7. Schematic showing amino acid sequence and probable secondary structure producing 12 transmembrane domains of the human reduced folate carrier (RFC). Modified from Zhao *et al.*, 1999; Drori *et al.*, 2000a; Cao & Matherly, 2004).



The terminal regions are not thought to play a role in the protein's function (Wong *et al.,* 1995; Marchant *et al.*, 2002; Flatley *et al.*, 2004), but like many proteins they contain motifs recognized for membrane localisation and intracellular trafficking (Lodish *et al.*, 1995).

RFC maintains a stable tertiary structure using, amongst other bonds, charged pair attraction between negatively charged aspartate residues at 88 and 453 interacting with a positively charged arginine at 133 (TMD 4) of human RFC (Liu & Matherly., 2001). Similarly proline residues in the α-helical TMD are important features for RFC to adopt its tertiary structure (Brigle *et al.*, 1995). It is therefore not surprising that the introduction of more proline, or substitution of existing proline will produce kinks and alter the protein structure and function (Wong *et al.*, 1999).

With the exception of the short sequence K_{204}RPKRSLFFNR_{214} that makes up the first 11 residues of the intracellular loop, the majority of residues in the loop are not conserved and their substitution had minimal effect (Liu *et al.*, 2003; Witt *et al.*, 2004). However, the overall size of the large intracellular loop does appear to be crucial in development of the correct protein conformation (Sadlish *et al.,* 2002a; Liu *et al.,* 2003). These findings support the theory proposed by Sadlish *et al.,* (2002a) that only 6 TDMs can be inserted into the cell membrane at one time, the loop keeps them far enough apart for two insertions but maintains the proximity of each to then form the functional protein.

Glycosylation at residue 58 (asparagine) accounts for the difference in molecular weight (Matherly and Angeles, 1994), but its role is poorly understood. Wong *et al.,* (1998) reported a slight loss in RFC activity when a mutant non-glycosylated protein was studied. Nehls *et al.,* (2000) proposed that glycosylation is required for intra-

cellular localisation and trafficking of most proteins through the cell, but earlier work by Loo and Clarke (1994) was unable to show a difference between the amount of wild-type or mutant RFC getting to the membrane. They concluded that glycosylation of RFC played no role in its function or its ability to migrate to the cell membrane and establish itself as a transmembrane structure. They did however demonstrate an increased turnover of the mutant RFC and suggest that the loss of activity is due to the influence that glycosylation has on the half-life of the protein.

1.7.2 THE RFC-1 GENE

Once the protein structure was identified it was relatively easy to generate cDNA for the protein and find a match in the human genome. A sequence on chromosome 21 at SLC19A1 was identified and confirmed using fluorescent *in-situ* hybridisation (FISH). Further refinement located the gene coding for the human RFC protein to the long arm of chromosome 21, specifically the 21q22.2-q22.3 region (Moscow *et al.*, 1995; Yang-Feng *et al.*, 1995; Lapenta *et al.*, 1998).

The structure of the gene has been studied in mouse, hamster and human cell lines and is referred to as RFC-1. In humans, Tolner *et al.*, (1998) studied cDNA from 16 genomic clones and obtained 2 overlapping sequences covering the RFC-1 gene. The first strand, labelled λhRFC1-1, is a 19kb segment containing exons 1, 2, 3 and 4. The second strand, λh-RFC1-2 is slightly shorter at 17kb and contains exons 5 and 6. The 5′ end of λhRFC1-2 overlaps 1.5kb of the 3′ end of λhRFC1-1 and thus together cover 34.5 kb. The full-length cDNA was confirmed as the RFC-1 gene by transfection into a methotrexate resistant cell line such as MTX^r Chinese Hamster Ovary cells (Trippet & Bertino, 1999; Wong *et al.*, 1995; Moscow *et al.*, 1995). Such cells have become resistant to this folate analogue by acquisition of a mutation in the

protein that renders it non-functional and unable to take-up MTX. When the proposed gene was transfected into these cells and they started to transport MTX across the cell membrane, (followed using ^{3}H-MTX), then this was strong evidence that the genetic material introduced does indeed code for RFC.

The exon sequences have been submitted to both the National Center for Biotechnology Information (NCBI) and European Molecular Biology Laboratory (EMBL) databases, using the following accession numbers (Table 1.1).

Table 1.1. Accession numbers for RFC-1 exons on the NCBI Entrez (*http://www.ncbi.nlm.nih.gov/Entrez/*) and EMBL (*http://www.ebi.ac.uk/embl/index.html*) databases.

Exon	Accession No.	Size (bp)
1a, 1b and 1c	U92868	3772
2	U92869	250
3	U92870	772
4	U92871	214
5	U92872	151
6	U92873	1451

Following Rapid Amplification of cDNA Ends (RACE) and gene walking studies, the intronic regions between the exons have been defined and show several intron-exon-intron splice sites that comform to the GT-AC rule. A shorter cDNA sequence was submitted to GenBank (accession number U19870) by Matherly's group (Wong *et al.*, 1995) containing a 1776 bp open reading frame (ORF). The cDNA ORF from either source predicts a protein of 64,873 Dalton which is in the range for the deglycoslyated protein obtained experimentally and demonstrated by Western blot (Wong *et al.*, 1998).

As can be seen from table 1.1, Tolner *et al.,* (1998) describe three alternative untranslated sequences in exon 1, labelled 1a, 1b and 1c. Each of them ends with the AG sequence that provides the opportunity for alternative splicing first postulated by Williams *et al.,* (1994) and produce three RFC-1 variants, of which variant II produces the most efficient transcription.

- Variant I contains exons 1a plus exons 2-6
- Variant II contains exons 1b plus exons 2-6
- Variant III contains exons 1c plus 2 and part of 3

Variants I and II are active proteins, but the truncated RFC from variant III is inactive. Using different cell lines, Gong *et al.,* (1999) identified a fourth alternative, exon 1d, from foetal liver cells and showed that the alternative splicing appeared to be tissue specific. Such tissue specificity was later confirmed by Whetstine *et al.,* (2002). Exons 1a and 1d are separated from their neighbours by large intronic regions and appear to have independent promoters. However, with only 21 bases between exons 1b and 1c they are likely to share the same promoter upstream of exon 1b.

Seven possible non-coding exons containing either promoters or promoter enhancers have now been identified up to 35kb upstream and can result in 18 unique splice variants (Whetstine *et al.*, 2002b). The role of each of these proposed variants has yet to be completely understood, but some are tissue specific and respond to specific transcription factors (Matherly & Goldman, 2003; Peyton *et al*, 2007). One of the newly defined non-coding exons (Flatley *et al.,* 2004) introduces an alternative start codon. This alternative start continues to produce a functional protein because the effect of the splicing is to simply add 64 residues to the N-terminal without shifting the triplet sequence of the rest of the protein. This larger alternative has higher K_M

(indicating that binding has been reduced), and also has a lower V_{max} (reduced activity). It is of particular interest here because this splice variant is a major form in ALL (Flatley *et al.*, 2004), but no possible mechanism was suggested or obvious.

Although there is increased RFC-1 expression in response to very low folate concentrations (Sierra *et al.,* 1999), there is little known about regulation of RFC-1 expression. When 5-formyltetrahdyrofolate is reintroduced to 'starved' cells, the cells revert to normal transcription rates, suggesting that a feedback mechanism between intracellular folate levels and RFC transcription must exist (Jansen *et al.,* 1990). Gong *et al.,* (1999) measured mRNA levels throughout the normal cell cycle and noted that transcription is linked to cellular requirements for folate, reaching a peak during the G1/S phase. In fact, they showed that the variable transcripts show a chronological pattern during the cell cycle, with exon 1c being expressed earlier than 1b and 1d. The significance of this is still unknown.

To confuse exon identification further, Matherly's group (Zhang *et al.*, 1998a) used a different numbering system for the exons. They placed the Kozak start sequence (ATG) in exon 3, ie., corresponds to exon 2 in Tolner's model (Tolner *et al.*,1997). In a series of experiments to determine the RFC-1 gene promoters, Zhang *et al.,* (1998a) studied a 342 bp intronic region between exon 2 and 3 (1b and 2 in Tolner's model) which they labelled Pro32, and a further 996 bp region upstream of exon 1 (1a in Tolner's model) which was labelled Pro43. These regions were later reclassified as hRFC-A and hRFC-B respectively (Whetstine & Matherly, 2001a). The hRFC-B promoter is more sensitive to suppression by the p53 protein (Matherly, 2001). This provides a mechanism by which the cell can reduce the concentration of intracellular folates necessary for cell replications and contribute to the p53 inhibition of damaged

DNA (Lohrum & Vousden, 2000; Gross *et al.*, 2001; Ding *et al.*, 2001a). Further discussion on factors affecting these promoter regions are reviewed in Bosson, 2003.

Whilst a lot of attention has been devoted to the 5′-UTR it appears that exon 6 at the other end may be just as variable and can explain some of the differences seen for the cytoplasmic C-terminal between species (Brigle *et al.*, 1997; Zhang *et al.*, 1998b). Further control of cellular folate uptake, at the level of transcription, is likely to occur as a result of DNA methylation. This was shown to occur in mice (Jansen *et al.*, 1998) and is also likely to occur in humans as it does for so many other genes.

1.7.3 RFC AND METHOTREXATE RESISTANCE

In section 1.6 it was highlighted that the most effective mechanism the cell can adopt for MTX resistance is to prevent it entering the cell in the first place. Whilst at very high concentrations some will diffuse across the cell membrane along a concentration gradient, the intracellular MTX concentration is essentially dependent on the RFC. Inhibited cellular uptake can be achieved in one of four ways.

1.7.3.1 **Down-regulation of the RFC-1 gene**

The simplest and most responsive way to reduce the amount of RFC in the cell membrane is to suppress the gene coding for the protein (Zhang *et al.,* 1998c). $MTX^{R}ZR$-75-1 and K500E are both MTX resistant cell lines which have mRNA levels significantly lower than those found in the corresponding wild-type cells and utilises this mechanism (Moscow *et al.,* 1995; Ding *et al.,* 2001b). Simple down-regulation of the RFC-1 gene (Rothem *et al.*, 2003), possibly by a mutation in transcriptional regulation, chromosomal translocation that produces transcriptional silencing (Ding *et al.,* 2001a) or by altering the methylation status (Worm, 2001; Rothem *et al.*, 2003) are all accepted ways in which this can be achieved. This suppression could be specific to particular spliced mRNA arrangements; eg., exon 1c was found to be the most frequently suppressed in studies by Gong and fellow workers (1999). The work by Ding *et al.,* (2001a) into the control of RFC-1 expression by the p53 gene opens-up the possibility that MTX resistance can be indirectly achieved by a mutation in the p53 gene.

Gorlick *et al.,* (1997) reported decreased expression in a series of relapse ALL cases, indicating that this mechanism is used *in-vivo* and not just a result of 'forced' resistance in cell lines grown *in vitro*. With this information, measurement of RFC-1

mRNA is not just a research tool, it could also be used in real-time PCR to identify those patients who will not respond to standard dose MTX based treatment and allow more effective treatment to be initiated.

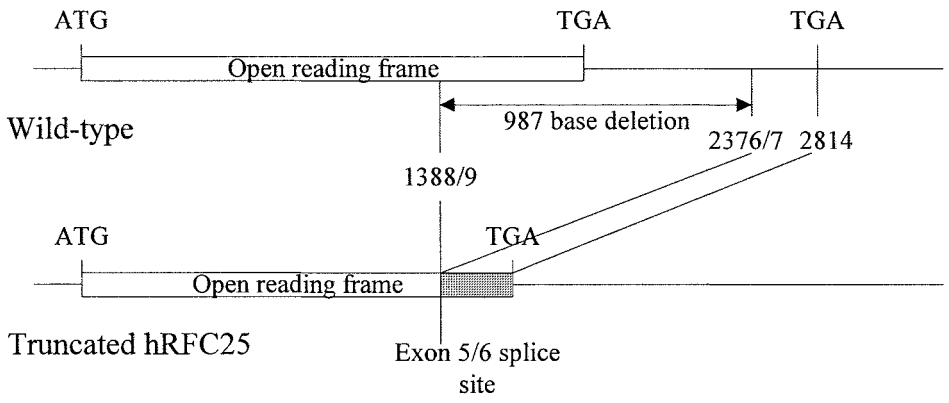
1.7.3.2 **Gross modification of the RFC protein structure**

In resistant cell lines where the mRNA is normal or even increased (Jansen *et al.,* 1990) the explanation must be a modified protein. Such modifications can be small (eg., a single amino acid) or gross changes in the structure as a whole. What has been noted is that changes have tended to cluster into specific regions of the protein (Rothem *et al.,* 2002).

A 7 bp deletion at the intron/exon 6 splice site (ie., nucleotides 1152 to 1158) in the murine MtxRIIOua^{R}2-4 cell-line, causes the loss of 2 residues and a frame shift (Sadlish *et al.,* 2000). The amino acid sequence resulting from the frame shift has no homology with the original and a premature stop codon is introduced. The resulting truncated protein with a modified C-terminal is less stable and whilst normal RFC in the membrane has a turnover of approximately 50% per 24hrs, the defective protein is cleared from the membrane faster than normal RFC (Wong *et al.,* 1999; Sadlish *et al.,* 2000). In humans, similar truncating mutations have been identified in MTX resistant sublines of MOLT-3 (Gong *et al.,* 1997; Yang *et al.*, 2003) and CCRF-CEM (Gifford *et al.*, 1998; Wong *et al.,* 1999; Yang *et al.*, 2003). MOLT-3/MTX$_{10,000}$ cell line contains 2 truncating mutations (Gong *et al.,* 1997). When both mutations occur together, one in each allele, the resistance is high (ie 10,000 fold). This and other early termination mutations are reviewed in Bosson (2003) as are techniques used to investigate the putative protein structure.

Not all truncated RFCs produce MTX resistance. Drori *et al.*, (2000a) described a 987 base deletion in the CEM-7A leukaemic cell line which causes the loss of 160 normally coded amino acids from the C-terminal and the normal stop codon, (figure 1.8). The deletion of bases 1389 to 2376 produce an alternative open reading frame with 58 alternate C-terminal residues before reaching a new stop codon in the previously untranslated 3′-region. The proposed structure of the resulting RFC lacks the last transmembrane domain causing the C-terminal to reside outside the cell membrane (Drori *et al.*, 2000a). Despite Drori *et al.*, (2000a) proposing that overexpression of this 'alternatively spliced gene' could still transport sufficient MTX to make it sensitive, Matherly (2001) considered it nonfunctional.

Figure 1.8. Alternative splicing of cDNA from human leukaemic CEM-7A cells producing a truncated RFC (Drori *et al.*, 2000a



1.7.3.3 **Impaired transport function**

In order for RFC to transport folate across the membrane it is predicted that the protein will undergo a conformational change. Such a mechanism will require several of the TMDs to interact with intracellular domains (Zhao *et al.*, 1998a). For a long time the intracellular loop sequences connecting the TMDs were thought to be just

that, connectors, but when the primary structure was modified by the insertion of an immunogenic marker sequence, the RFC activity was modified (Ferguson & Flintoff, 1999; Zhao *et al.,* 1998a; 2003).

Forced mutations of Arg133 and Asp88 in human RFC by Liu and Matherly (2001) provide further evidence that the TMDs interact with each other to produce a functional carrier. The two oppositely charged amino acids form an ion-pair. When both are mutated to neutral amino acids they can still form an association and have minimal effect on activity, but if only one of the pair is changed, it prevents the tertiary structure being formed and results in a total loss of activity. Similar experiments by Sadlish *et al.,* (2002b) on Arg373 in murine cells indicated the charge of this residue was just as important and Flintoff *et al.,* (2004) found that Gly259 in the large cytoplasmic loop and Leu291 in the extracellular loop between TMDs 7 and 8, cause decreased RFC activity. In this latter case, the data indicates Gly259 is not involved in folate binding and therefore must play a part in the proposed conformational change to transport folate across the membrane.

1.7.3.4 **Modified binding**

The most effective way for the cell to obtain sufficient reduced folate to sustain a high replication rate and yet still exclude MTX from the cell is to make the RFC selective. This can be achieved by substituting amino acid residues responsible for binding recognition or affinity and is demonstrated experimentally by changes in K_M.

Amino acids 45-48 appear to form an aqueous pocket in the first TDM that would be important for the binding of folates and MTX. It is not surprising therefore, that mutations in the first TMD have an effect on affinity, some directly and others indirectly (Drori *et al.*, 2000b; Gifford *et al.*, 2002; Flintoff *et al.*, 2003). The results

of targeted mutation studies performed by Zhao and co-workers (2000) show that substitution of the negatively charged glutamic acid at position 45 with another negatively charged amino acid, ie., aspartic acid, significantly decreases affinity for all four compounds. This indicates that it is not the negative charge at position 45 that is essential for RFC function. If substituted by a hydrophobic amino acid (eg., leucine or tryptophan), then affinity is also decreased, but substitution with hydrophilic glutamine provides decreased affinity for MTX whilst increasing affinity for folates. The benefit of a hydrophilic amino acid gives support to the aqueous pocket theory.

The effect of other substitutions investigated by Tse *et al.,* (1998), Zhao *et al.,* (1998a, 1998b, 2000), Drori *et al.*, (2000b), Sharina *et al.,* (2001), Witt & Matherly, (2002), Yang *et al.*, (2003) have been reviewed by Bosson (2003) and further described by Flintoff *et al.*, (2004).

1.8 FOLYLPOLYGLUTAMATE SYNTHETASE (FPGS) and 1.9 FOLYLPOLYGLUTAMATE HYDROLASE (FPGH)

Antifolates, like natural folates are retained within the cell and are more active when polyglutamylated (Jolivet *et al.*, 1983, Kim and Shane, 1994). Using HPLC it is clear that the intracellular polyglutamylated status of folates and antifolates within the cell is partially dependent on the relative activities of enzymes involved in the polyglut-amylation process play a significant role (Mauritz *et al.*, 2002; Cheng *et al.*, 2006).

Folylpolyglutamate synthetase (FPGS; tetrahydrofolate:L-glutamate-γ-ligase (ADP forming); EC 6.3.2.17) catalyses the addition of glutamate residues to folates and antifolates, and thus enhance cellular retention (Chen *et al.*, 1996). Cells which lack FPGS activity present with a degree of antifolate resistance (McCloskey *et al.*, 1991; Roy *et al.*, 1995; Zeng *et al.*, 2001; Liani *et al.*, 2003). FPGS activity is lower in T-ALL and may account for the poorer prognosis of T-ALL (Rots *et al.*, 1999; Rots *et al.*, 2000b).

Folylpolyglutamate hydrolase (EC 3.4.19.9; also known as γ-glutamyl hydrolase; GGH or γGH) catalyes the hydrolysis of glutamate residues from polyglutamylated folates and antifolate, facilitating efflux from the cell via the MRP mechanism (Cheng *et al.*, 2006). Increasing the activity of this enzyme will reduce the amount of polyglutamylated MTX and reduce its pharmacological action. Whilst Zheng *et al.*, (2001) suggest that hydrolase activity is not as important as FPGS in determining methotrexate efflux, others have supported consideration of its activity as a ratio; FPGS against FPGH (Rots *et al.*, 1999) or polyglutamylated MTX concentration against FPGH activity (Chave *et al.*, 2002) as potential predictors of MTX efficacy.

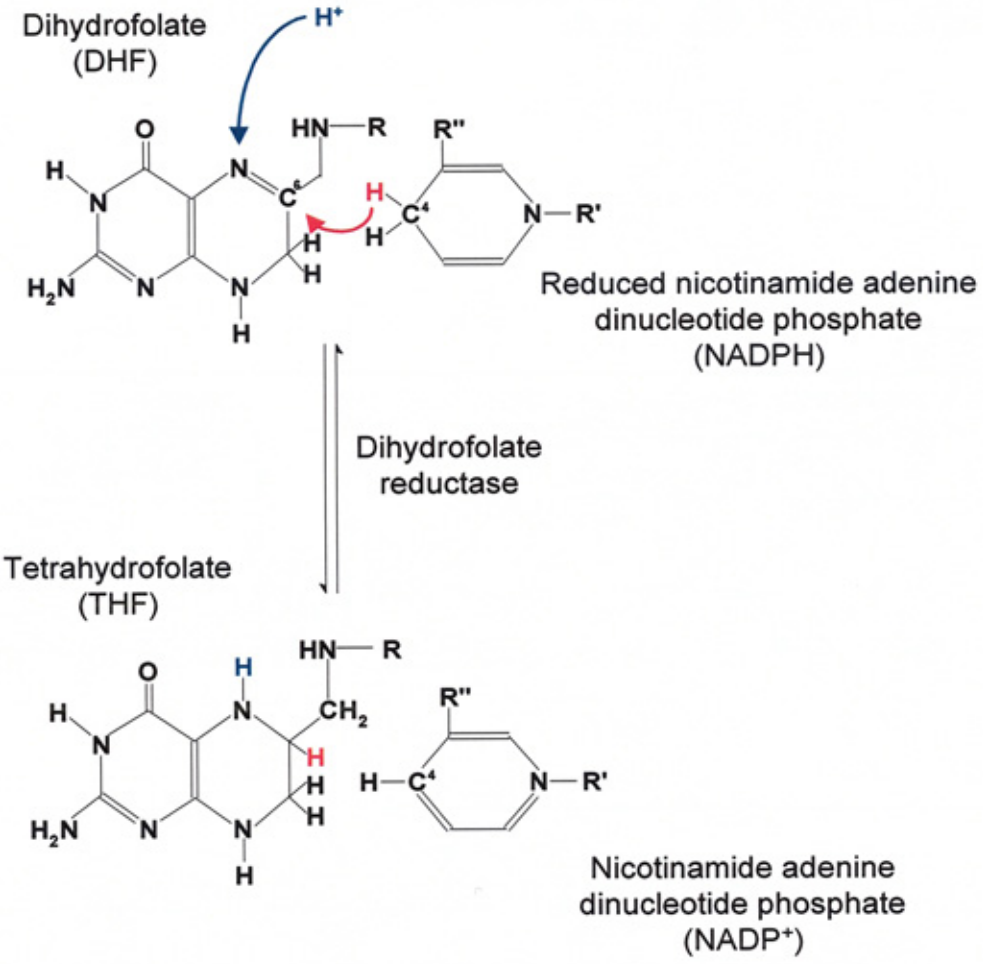
1.10 DIHYDROFOLATE REDUCTASE (DHFR)

1.10.1 ENZYME STRUCTURE AND FUNCTION

Only a small amount of folate is required to maintain cellular 1-carbon metabolism due to the efficient regeneration of DHF back to THF by dihydrofolate reductase (DHFR; 5,6,7,8-tetrahydrofolate:$NADP^+$ oxidoreductase; EC 1.5.1.3) (O'Neil & McKusick, 2000) (figures 1.3 and 1.9). This key enzyme is therefore the primary target of many antifolate drugs, including MTX.

Transfer of the hydrogen plus its electron pair (hydride ion, $:H^-$) from carbon-4 of NADPH to carbon-6 of the DHF provides one of the hydrogens, and at the same time a proton (H^+) is added to nitrogen-5 (Figure 1.9) to provide the other. The protonation step is naturally dependent on pH and is facilitated by a simple change in conformation of the enzyme that modifies the pKa value of an aspartate residue (Rod and Brooks, 2003). For successful transfer of the hydride ion, the donor and acceptor atoms must be at 180° to each other and with an optimum distance between them (Rajagopalan & Benkovic, 2002; Thorpe & Brooks, 2004). Therefore, like many enzymes, the amino acid sequence and tertiary structure adopted by them is crucial and even slight changes could have significant effects. Catalysis is achieved by multiple microvibrations along a plane of amino acid R-groups which lowers the overall activation energy required for hydride transfer (Rajagopalan & Benkovic, 2002; Watney *et al.*, 2003; Schnell *et al.*, 2004). Blakley *et al.*, (1993) determined the dissociation constants for each of the steps in the reaction and have shown that under acidic and neutral pH conditions the rate-limiting step is the dissociation of THF from the binding site. At higher pH it is the lack of H^+ that becomes the limiting factor for catalysis.

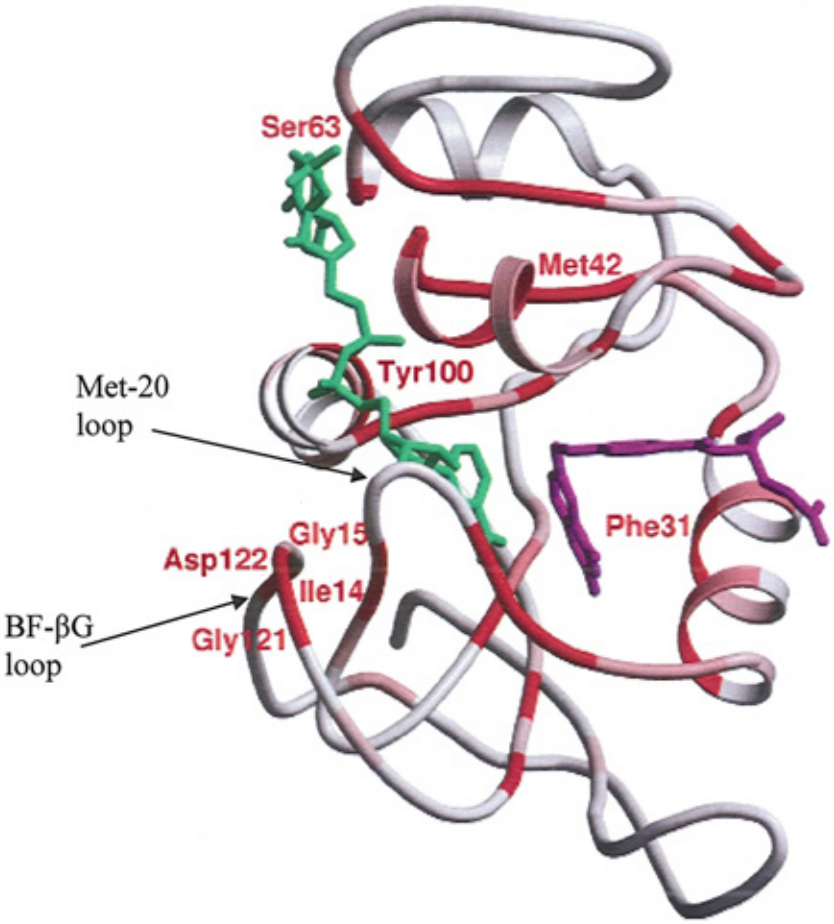
Figure 1.9. Dihydrofolate reductase catalysed transfer of hydride (shown in red) from NADPH and addition of proton (shown in blue). R = PABA-glutamate residue of folate; R' = ribose phosphate group; R'' = amide group. (modified from Voet *et al.*, 2006)



For such an important enzyme, it is a relatively small protein, with a mass of 21,544 Daltons, made up of only 186 residues (Blakley & Sorrentino, 1998) which are highly conserved across the species (Bullerjahn & Freisheim, 1992) (Figure 1.10). The bigger 44 kDa protein identified in humans by Rothenberg & Jabal (1982) turned out to be simply a homodimer. In *Escherichia coli* it exists as an even bigger homotetramer (Schmitzer *et al.*, 2004). The solubility of the enzyme makes it suitable for x-ray diffraction studies, with and without substrate bound. This has allowed

elucidation of the 3-D configuration and appreciation of what happens during binding and catalysis (Stockman *et al.*, 1992).

Figure 1.10. Ribbon representation of DHFR structure showing conserved sequences and specific amino acids across species in red; NADPH shown in green; DHF shown in purple. (Taken from Agarwal *et al.*, 2002).



The structure has a deep groove made up of hydrophobic amino acids, which along one face form the binding site for DHF and a second binding site for NADPH (Stockman *et al.*, 1992). The pteridine ring of the folate fits into the bottom of the groove with the glutamyl residues lying along the protein surface. NADPH is positioned so that its ring fits into the hydrophobic cleft alongside the folate. Because of the flipped pteridine ring of MTX it forms a tighter bond with the residues in the active site and inhibits activity by increasing the dissociation time (Belur *et al.*, 2001).

When either DHF or NADPH bind to the enzyme it undergoes a slight conformational change that gives the binding site a greater affinity for the other substrate molecule (Bullerjahn & Freisheim, 1992).

The protein consists of seven parallel β-sheets and four α-helices connected by loops (figure 1.10). Three of the loops, MET-20 (residues 14-24), F-G loop and G-H loop, form important structures that provide the flexibility for the enzyme to adopt the open, closed or occluded conformations (Bullerjahn & Freisheim, 1992; Rod & Brooks, 2003; Schnell *et al.*, 2004). In the occluded position methionine at position 16 occludes the NADPH binding site, whilst in the closed position it is moved out of the binding site and allows the terminal hydroxyl of NADPH-ribose moiety to form a hydrogen bond with alanine at position 19 (Thorpe & Brooks, 2004). The full reaction and conformational changes that must occur have effectively been shown in an animated clip created by Sawaya and Kraut, available on the University of California San Diego (UCSD) website, (*http://chem-faculty.ucsd.edu/kraut/dhfr.html*).

The flexibility of the protein, especially the loops and the exact positioning of the DHFR and NADPH molecules make the primary structure and resulting topology of the enzyme very important. As already mentioned, any mutations are likely to have a significant effect on enzyme activity. This is thought to be the reason for seeing so few mutations or differences between species, as even subtle changes can be lethal to the cell (Tan *et al.*, 1990; Stockman *et al.*, 1992; Ivernizzi *et al.*, 2003).

Iqbal *et al.,* (2000) reported *in vivo* synthesis of DHFR isoforms when patients with a range of haemopoetic cancers were exposed to colony stimulating factors, although the significance of these isoforms has still to be elucidated.

1.10.2 THE DHFR GENE

In humans the functional gene has been located to a 30 kb region on chromosome 5 at 5q11.2-q13.2. (Anagnou *et al.*, 1984; Chen M-J *et al.*, 1984; Funanage V L *et al.*, 1984; Maurer *et al.*, 1984). Originally described by these authors to have 6 exons and recorded on the NCBI database (X00855) with the start codon in exon 1. However, Ensembl (ENST00000307796) and SwissProt (P00374) databases use a 7 exon numbering system with the start codon in exon 2 (the 7 exon numbering system will be used in this thesis). The splice sites for intron-exon boundaries follow the Chambon rule (5'-GT and AG-3') and are highly conserved across the species (Chen M-J *et al.*, 1984) with homology across the whole of the gene being 84% concordant between human and mouse DNA (Maurer *et al.*, 1984).

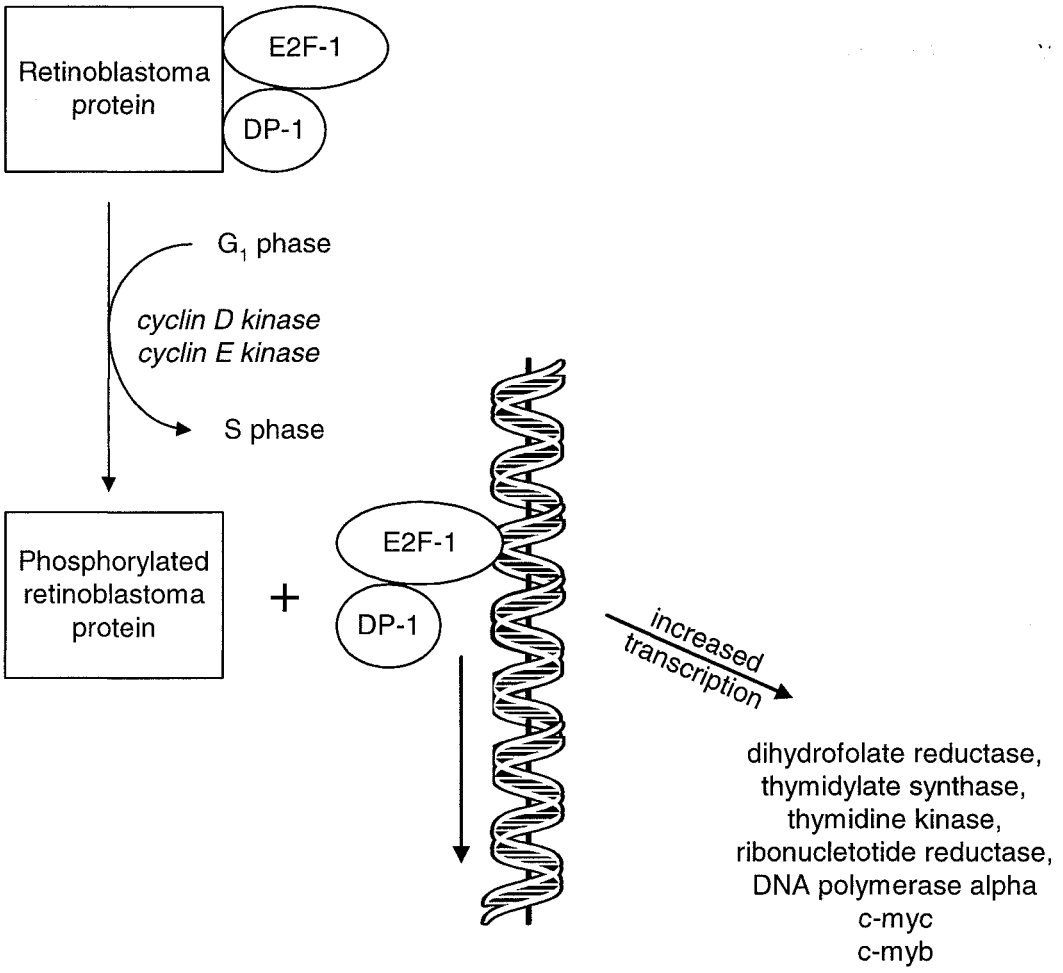
Two years before the gene was fully described, Chen *et al.*, (1982) and Srimatkandada *et al.*, (1983) had reported extra-chromosomal double-minutes in the nucleus that would provide extra copies of the gene (Pauletti *et al.*, 1990; Singer *et al.*, 2000). Singer *et al.*, (2000) proposed that defects in p53 that proceed malignant changes will tolerate chromosomal rearrangements and breakages of repeat bridges that release this extra-chromosomal material. Concurrent homogenously staining regions (HSR) on chromosomes 3, 6 and 18 are now known to be intronless, non-transcribed pseudogenes (hDHFR$\psi_{1\text{-}4}$) (Anagnou *et al.*, 1984; Chen M-J *et al.*, 1984; Funanage V L *et al.*, 1984; Maurer *et al.*, 1984). These evolutionary artefacts have most likely been generated as double-minutes synthesised during periods of need, but which have then been reincorporated back into the genome on the back of viral reverse transcription (Anagnou *et al.*, 1984). Differences in racial distribution suggest that pseudogenes are relatively recent evolutionary events (Chen *et al.*, 1982).

1.10.3 REGULATION OF THE DHFR GENE AND mRNA TRANSLATION

Transcription of the DHFR gene is primarily under control of the E2F transcription factor which facilitates an increase in the enzyme as the cell moves from G_1 to S-phase of the cell cycle and the demand for nucleic acids is greatest (Banerjee *et al.*, 2002; Sowers *et al.*, 2003). The relationship between DHFR, E2F, retinoblastoma gene and P53 is shown in figure 1.11 (Gorlick *et al.*, 1996). Normally E2F, and an associated protein (DP-1), are bound to unphosphorylated retinoblastoma protein (Rb). As the cell ends G_1 phase, cyclin D and E kinases phosphorylate Rb and this pushes off the E2F protein so that it can now bind to the 5'-TTT(C/G)(C/G)CGC-3'promoter of the DHFR gene (Wade *et al.*, 1995). Loss of retinoblastoma suppression is seen in several malignancies and thus facilitates the replication of malignant cells (Banerjee *et al.*, 2002; Serra *et al.*, 2004). Goto *et al.,* (2001) indicate that the 3'-UTR region of the gene is responsible for polyadenylation of the transcribed mRNA and that this confers greater stability and thus a greater chance to undergo translation.

To provide finer control of DHFR activity there is evidence that DHFR binds and inhibits its own mRNA (Chu *et al.*, 1993; Tai *et al.*, 2004; Skacel *et al.*, 2005). Using gel-shift experiments, Tai *et al.,* (2004) identified that bases 401-482 of human DHFR mRNA recognise and bind inactive enzyme; probably involving Cys-6, Ile-7, Arg-28 and Phe-34 (Tai *et al.*, 2002). When intracellular levels of DHF are high, the conformational change of the enzyme to the open configuration reduces this affinity to inhibit its mRNA. The permitted translation ultimately increases total enzyme activity to regenerate DHF to THF. In true negative feedback fashion, as the DHF levels fall the mRNA will again be inhibited so that activity matches need.

Figure 1.11. Proposed mechanism for retinoblastoma and E2F control of key enzymes (including DHFR) required for cell replication. Modified from Gorlick *et al* (1996).



1.10.4 METHOTREXATE RESISTANCE AND DHFR

The flipped pteridine ring of MTX has a strong affinity for DHFR generated by an interaction between it and the R-group of glutamic acid at residue 30 (Blakley *et al.*, 1993). True folates cannot make this bond and this goes some way to explain the comparatively higher affinity that makes the drug such an effective competitive inhibitor. MTX resistance at this point will therefore require a change to the amino acid sequence that affects the binding affinity; or an increase in the amount of enzyme

present to overcome the inhibitor. Whilst a few cell-lines have been produced *in vitro* with altered affinity, none have been noted *in vivo* (Spencer *et al.*, 1996). This leaves the latter mechanism as that which explains the inherent resistance seen in some cancers (Cowan *et al.*, 1982; Serra *et al.*, 2004), racial and ALL sub-groups (Matherly & Taub, 1999; Rots *et al.*, 2000c; Levy *et al.*, 2003) that correlates with relapse.

Resistance noted in the early days of MTX use (Bertino *et al.*, 1963) was attributed to an increase in DHFR levels. Whilst this increase is likely to influence MTX effect, the increase is now recognised to be mainly a simple response of the negative feedback system described earlier (section 1.10.3) confirmed by quantitative reverse transcription studies (Serra *et al.*, 2004). MTX binds to DHFR and causes the same conformational change that reduces the inhibition of DHFR mRNA (Eastman *et al.*, 1991; Chu *et al.*, 1993; Tai *et al.*, 2004; Skacel *et al.*, 2005). The resulting increase in DHFR is such a strong factor in indicating outcome, that Matherly *et al.*, (1997) proposed assessment of DHFR levels as one of the treatment stratification parameters, and yet this does not appear to be used in routine diagnostic procedures. This may be because of complicating polymorphisms influencing gene expression (Goto *et al.*, 2001; Mesner and Hamlin, 2005). Recently, the research group lead by Evans (Sorich *et al.*, 2008) highlight that the cells need to be in the S-phase in order for MTX to have a pharmacological effect, and that a low DHFR activity (ie., low cellular proliferation activity) also correlated with poor *in vivo* response.

1.10.5 EFFECTS OF FORCED MUTATIONS IN DHFR

Recognising that the levels of MTX used *in vitro* to generate resistant cell-lines is much higher than that encountered *in vivo* and therefore may not correlate with causes of relapse, they have still been invaluable in understanding how the enzyme works.

The effects of mutations on activity have been relatively easy to follow using the intrinsic fluorescence of NADPH ($270_{ex}/320_{em}$ nm) which is quenched when bound to the enzyme. A summary of the mutations studied are shown in table 1.2 and only key observations highlighted here. The cysteine residue at position 6 plays probably the most important role in binding between DHFR and its mRNA (Tai *et al.*, 2002). The substitution of this amino acid almost totally removes the negative feedback control and allows production of enough enzyme to bind any MTX present and still have enough uninhibited active enzyme to regenerate reduced folates and survive.

Significant reduction in affinity was noted when glycine at position 15 (Dicker *et al.*, 1993) or leucine at 22 (Cody *et al.*, 2005) were substituted. When tryptophan-24 is substituted the conflicting increase in activity at the same time as affinity is decreased is a result of the ease with which the product dissociates (Beard *et al., 1991,* Huang *et al.*, 1989). Affinity is also reduced when phenylalanine-31 is substituted (Prendergast *et al.*, 1989; Dicker *et al.*, 1989; Schweitzer *et al.*,1989; Srimatkandada *et al.*, 1989; Tsay *et al.*, 1990; Miyachi *et al.*, 1993;Chunduru *et al.*, 1994; Blakley and Sorrentino, 1998). In 1993, Blakley and Sorrentino added that this phenylalanine and that at residue 34 were important in holding a single water molecule necessary for hydride transfer.

The need for loop mobility to facilitate conformational changes during catalysis was demonstrated when resides 45 and 46 were removed (Tan *et al.*, 1990) or the C-terminal loop modified (Bullerjahn and Freisheim, 1992) as the resulting activity reduced by at least half. As well as mobility of the protein structure, distances and angles between residues and the substrates are equally critical (Agarwal *et al.*, 2002; Watney *et al.*, 2003).

Table 1.2. Summary of reported mutations in DHFR. (H = human; m = mouse; e = *Escherichia coli*)

Nucleotide affected	Amino acid reside affected	Species	Domain affected and effect on DHFR	Ref
$_{19}TGC_{21}\rightarrow GCT$	Cys6Ala	H	Loss of negative feedback on mRNA	Tai *et al.*, (2002)
$_{19}TGC_{21}\rightarrow AGA$	Cys6Ser	H	Loss of negative feedback on mRNA	Tai *et al.*, (2002)
C46T	Gly15Try	m	$\uparrow K_M$	Dicker *et al.*, (1993)
$_{67}CTG_{69}\rightarrow CGG$	Leu22Arg	m, H	>250 fold $\uparrow K_M$	Cody *et al.*, (2005), Ercikan-Abli *et al.*, (1996)
$_{67}CTG_{69}\rightarrow ATG$	Leu22Met	H	3 fold $\uparrow K_M$	
$_{67}CTG_{69}\rightarrow TTT$	Leu22Phe	H	2 fold $\uparrow K_M$	Lewis *et al.*, (1995)
$_{67}CTG_{69}\rightarrow AAT$	Leu22Ile	H	7 fold $\uparrow K_M$	Ercikan-Abli *et al.*,
$_{67}CTG_{69}\rightarrow TAC$	Leu22Tyr	H	2 fold $\uparrow K_M$	(1996)
$_{73}TGG_{76}\rightarrow TTT$	Try24Phe	H	$\uparrow K_M$ $\uparrow V_{max}$	Beard *et al.*, (1991), Huang *et al.*, (1989)
$_{94}TTC_{96}\rightarrow TTT$	Phe31Try	H	$\uparrow K_M$ $\downarrow V_{max.}$ Involved directly with	Chunduru *et al.*, (1994)
$_{94}TTC_{96}\rightarrow TTG$	Phe31Leu	H	folate/antifolate binding	Prendergast *et al.*, (1989)
T95C	Phe31Ser	H		Tsay *et al.*, (1990)
T106C	Phe34Ser	H	3 fold $\downarrow K_M$ for NADPH; 24 fold $\uparrow K_M$ for DHF; $\downarrow V_{max.}$ Substantial reconfiguration of tertiary structure	Schweitzer *et al.*, (1989)
$del_{136}GGTAAA_{141}$	delGly45Lys46	H	$\rightarrow K_M$ $\downarrow V_{max.}$ Reduces conformational loop mobility during catalyis	Tan *et al.*, (1990)

Continued.....

Table 1.2 Continued

Nucleotide affected	Amino acid reside affected	Species	Domain affected and effect on DHFR	Ref
A160C	Lys54Gln	H	↓NADPH binding	Huang *et al.*, (1990)
G209A	Arg70Lys	H	$\uparrow K_M \downarrow V_{max}$	Thompson & Freisheim (1991)
G362T (G389T)	Gly121Val (Gly130 in humans)	e	$\downarrow V_{max}$. Modifies βF-βG loop so that distance between NADPH and DHF is increased.	Watney *et al.*, 2003
$del_{499}GGT\text{-}GTC_{507}$	delGly164-Val169	H	Loss of loop between βG and βH causes total inactivity	Bullerjahn & Freisheim (1992)

1.10.6 DHFR POLYMORPHISM

Analysis of the NCBI SNP database shows a number of polymorphisms, but few have data on their penetration. Whilst these may be important, most are intronic and few have been investigated. Whilst 16 are found in exons 1 or 2 and a further 13 occur in exon 7, none are found in the coding regions of this gene. Recent research into folate metabolism (Stanislawska-Sachadyn *et al*, 2008) and risk factors for neural tube defects (Johnson *et al.*, 2004, 2005; Parle-McDermott *et al.*, 2007) has identified a frequent SNP deletion in the 5'UTR (WT/WT 28%, WT/del 55%, del/del 17%) that may influence folate metabolism sufficiently that it should be considered in MTX treatment. The 19-bp deletion removes a Sp1 transcription factor-binding site and causes decreased transcription and hence DHFR activity (Farnham and Means, 1990).

1.11 METHYLENETETRAHYDROFOLATE REDUCTASE (MTHFR)

Methylenetetrahydrofolate reductase (EC 1.5.1.20; N^5N^{10}-methylenetetrahydrofolate reductase [NADPH]; MTHFR) plays a pivotal role in determining whether activated folates are directed into the remethylation of homocysteine or shunted into thymidine and purine synthesis (Wiemels *et al.*, 2001). Changes in activity at this key point can modify the effect of SNPs or mutations elsewhere in the complex interrelationships surrounding folate metabolism, MTX efficacy (Urano *et al.*, 2003) and even thiopurine drugs (Marinaki *et al.*, 2003).

As can be seen from figures 1.4 and 1.5, MTHFR converts N^5N^{10}-methylene-tetrahydrofolate to N^5-methyltetrahydrofolate, which is the co-substrate for remethylation of homocysteine to methionine. Its tertiary structure is still not fully understood, but it has a regulatory domain, the usual substrate binding site and a flavin adenine dinucleotide (FAD) binding site (Sibani *et al.*, 2003). The regulatory

domain responds allosterically to the intracellular level of S-adenosylmethionine (SAM) and thus provides a degree of negative feedback control and redirection of homocysteine into transulphation (Selhub, 1999; Steenge *et al.*, 2002). FAD is essential because it acts as a transient carrier of electrons as they are transferred from NADH to N^5N^{10}-methylenetetrahydrofolate; the step required to break the covalent bond between the methylene carbon and N^{10} group of the pteroate.

The MTHFR gene is found close to the tip of chromosome one, at 1p36.3, and comprises 11 exons that code for a 77 kDa protein (Goyette *et al.*, 1998) but which is found as a 154 kDa homodimer (Sibani *et al.*, 2003). Severe deficiency of MTHFR activity causes hyperhomocysteinaemia and homocystinuria which are associated with a whole range of morbidity (Goyette *et al.*, 1998). The total loss of enzyme activity is rare, but not unknown ((Goyette *et al.*, 1994; Sibani *et al.*, 2003).

There are several SNPs identified in the MTHFR gene that cause a milder form of enzyme deficiency, of which C677T is the most common and therefore potentially most important in considering the effect on efficacy of antifolates. The residue change, Ala222Val, is not directly linked with folate binding or catalytic activity but appears to reduce the stability of the dimer so that it has a lower affinity for FAD (Goyette *et al.*, 1998; Yamada *et al.,* 2001). The decrease in activity to around 35-50% of normal is still adequate under normal circumstances, but when folate status is low it can cause a mild hyperhomocysteinaemia (Goyette *et al.*, 1998).

If folate status is low, then a reduction of MTHFR activity can influence the aetiology (Weimels *et al.*, 2001; Friso *et al.*, 2002; Sharp *et al.*, 2002; Yi *et al.*, 2002) and response to treatment of several cancers (Kim, 1999; Duthie, 1999; Pakakasama *et al.*,, 2007) including ALL (Skibola *et al.*, 1999; Thompson *et al.*, 2001).

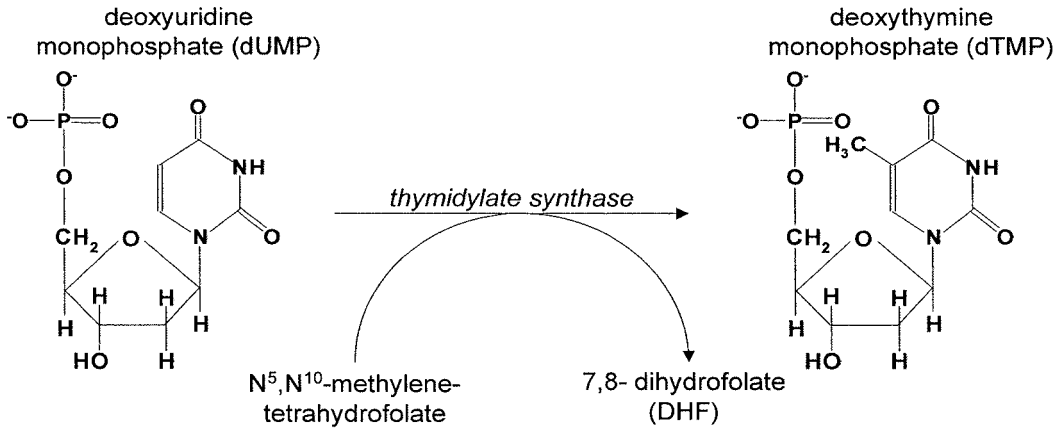
Another common SNP is the A1298C polymorphism. This causes a Glu429Ala substitution, that also produces a thermolabile enzyme, but this time the residue affected resides in the C-terminal regulatory domain and is therefore considered less significant (Sibani *et al.*, 2003). The polymorphism only becomes significant if the C1298 allele is found with the T allele of C677T (Weisberg *et al.,* 1998; de Jonge, 2003; Urano *et al.*, 2003) when de Jonge *et al* (2003) showed that TS activity was decreased.

With two independent SNPs it is difficult to determine what the wild type is, and the picture becomes even more complicated when all the proteins involved in folate metabolism are considered. Like several other proteins involved in folate metabolism, this gene can be alternatively spliced to produce a variety of mRNAs that reflect tissue specificity and cell cycle demand for enzyme activity (Goyette *et al.*, 1998).

1.12 THYMIDYLATE SYNTHASE (TS)

Cells entering, or already in, the G1/S phase of cell replication have a high demand for nucleic acids and this requires N^5N^{10}-methylenetetrahydrofolate to be diverted into purine and pyrimidine *de novo* synthesis. During these periods transcription factors stimulate increased levels of DHFR to facilitate rapid recycling of the relatively small folate pool. This is matched with the increased transcription of thymidylate synthase to ensure the deoxythymidine monophosphate (dTMP) pool is maintained (Sorich *et al.*, 2008). If dTMP levels fall, then there is an increased risk of misincorporation of other nucleic acids which are more readily available, (eg., uracil), with potentially fatal consequences for the cell (Goulian *et al.*, 1980; Wickramasinghe and Frida, 1994; Nijhout *et al.*, 2004). For this reason, the enzyme has become an important target for anticancer drugs.

Figure 1.12. Conversion of dUMP to dTMP by thymidylate synthase (TS) in the presence of methylenetetrahydrofolate as the carbon donor.



Thymidylate synthase (TS; methylenetetrahydrofolate:dUMP C-methyltransferase; dTMP synthase; TMP synthetase, TYMS; EC 2.1.1.45) catalyses the reductive methylation of deoxyuridine monophosphate (dUMP) to dTMP (figure 1.12). This 35.7 kDa protein is made up of 313 amino acids coded by 7 exons across a 30 kb section of chromosome 18 at 18p11.32 (Takeishi *et al.*, 1985). The 5'-untranslated region (UTR) has some unusual features that influence transcription of this enzyme and level of activity in the cell. The first is an unusually high GC content (approximately 80%); secondly there is the presence of variable number tandem repeats (VNTR) immediately upstream of the start codon, usually 2-3, but as many as 9 repeats have been reported (Mandola *et al.*, 2003). These tandem repeats have a common sequence which produces the third feature, stem-loop secondary structures of the mRNA. The VNTR produces two common versions of the 5'-UTR, identified as 2R or 3R (Kaneda *et al.*, 1987; Horie & Takeishi, 1997). The 3R tandem repeat results in a 2-4 times increase in the activity of TS compared to the 2R (Iacopetta *et al.*, 2001; Skibola *et al.*, 2002).

The anticancer drugs, 5-fluorouracil (5-FU) and MTX both inhibit TS (Banerjee *et al.*, 2002; Mauritz *et al.*, 2002), although the former is more effective on this enzyme. Resistance to 5-FU, MTX and some of the newer antifolates can be achieved by increased TS expression (Sigmond *et al.*, 2003). Like DHFR, fine control of TS expression is achieved by the ability of free enzyme to bind its own mRNA and suppress translation (Chu *et al.*, 1991). Hence Sigmond *et al.,* (2003) considered that the increase they were reporting was simply the feedback response expected when TS bound to drug (TS-drug complex) is no longer able to bind its own mRNA.

Iacopetta *et al.,* (2001) recognised that those patients who were homozygous for the triple repeat would not respond to 5-FU treatment as well as homozygote double repeats or heterozygotes. In the same way, the 3R/3R homozygotes render MTX less effective (Krajinovic *et al.*, 2002). Such a polymorphism would also afford some protection against misincorporation by ensuring relatively low levels of uracil (Skibola *et al.*, 2002). It is in this context that TS VNTR status and a G/C SNP in the second element at -58 (Mandola *et al.,* 2003; Kawakami and Watanabe G, 2003) is increasingly included in a pre-treatment screening profile of SNP's. By themselves the SNPs may have limited impact, but could have synergistic or counteractive effects in leukaemic or other cancer patients (Iacopetta *et al.*, 2001; Ezzeldin *et al.*, 2004).

1.13 OTHER KEY ENZYMES

Purine biosynthesis involves two folate dependent enzymes which are also inhibited by MTX and thus further potentiate the anticancer role of the drug. The first of these is glycinamide ribonucleotide transformylase (GART; EC 6.3.4.13) which catalyses the conversion of glycinamide ribotide to formylglycinamide ribotide using N^{10}-formyltetrahydrofolate as the carbon donor (Daubner *et al.*, 1986; Schild *et al.*, 1990; Brodsky *et al.*, 1997). A SNP in the gene for this enzyme is unlikely to confer resistance to MTX because the drug would still cause cell death via its effect on DHFR and TS. The second enzyme in the pathway susceptible to MTX inhibition is the penultimate reaction involving 5-aminoimidazole-4-carboxamide ribonucleotide formyl-transferase (AICART; EC 2.1.2.3). Like GART, it is unlikely that changes in this enzyme will confer resistance to MTX when they exist alone, but may enhance any effects elsewhere in the pathways involving folate/antifolates. Whilst DHFR, TS, GART and AICART are directly inhibited by MTX there are several other enzymes that, because of their involvement in folate metabolism, may indirectly influence the response to MTX and need to be briefly mentioned.

Serine hydroxymethyl transferase (SHMT; EC 2.1.2.1, also known as glycine hydroxylmethyltransferase) fails to get the attention that others have (Renwick *et al.*, 1998; Rao *et al.*, 2000), and yet it is crucial for the methylene transfer to folate that links DHFR and TS in the first place (figures 1.4 and 1.5) (Herbig *et al.*, 2002; Skibola *et al.*, 2004). The C1420T SNP reported in this enzyme by Skibola *et al.*, (2002) is associated with decreased folate levels, but needs further investigation to understand what the indirect effect on MTX efficacy are and how it can be used as a future target for chemotherapy (Rao *et al.*, 2000). The genotype of this gene was

shown to influences TS activity and thus effectiveness of MTX (de Jonge *et al.*, 2005).

Methionine synthase (MS or MTR; EC 2.1.1.13) catalyses the remethylation of homocysteine to methionine. Failure of this reaction is linked to a whole range of morbidity caused by the resulting hyperhomocysteinaemia (Goyette *et al.*, 1998) or lack of methionine (Friso *et al.*, 2002). The enzyme requires vitamin B_{12} (cobalamin) as coenzyme and deficiency will result in a drop in enzyme activity and hyperhomocysteinaemia. Loss of MS activity due to B_{12} deficiency, or a defect in the enzyme itself, will have an effect on folate metabolism because the relatively small amount of intracellar folate is no longer recycled, but gets stuck as 5-methyltetrahydrofolate. This is often referred to as the 'methyl trap' (Nijhout *et al.*, 2004) and yet, as the folate pathway '*backs up*', it is proposed that eventually $N^{5,10}$methylenetetrahydrofolate is pushed down the purine and pyrimidine biosynthetic pathways and confers a degree of protection against ALL (Skibola *et al.*, 2002). This enzyme is maintained in a reduced state by methionine synthase reductase (MTRR; EC 2.1.1.135). Defects in MTRR will present as defective MS (Leclerc *et al.*, 1998).

Under normal circumstances about half of the intracellular homocysteine is remethylated back to methionine for use in protein synthesis and as a methyl donor, but the remainder is irreversibly converted to cystathionine and then cysteine by cystathionine β-synthase (CBS; EC 4.2.1.22). This second pathway is referred to as the 'transulphation pathway' and controls *S*-adenosyl methionine (SAM) levels (Fillon-Emery *et al.*, 2004; Taub and Ge, 2005). This is important because SAM is a known allosteric inhibitor of MTHFR and can draw folates away from pyrimidine metabolism (Ito *et al.*, 1999; Taub and Ge, 2005). In the event that the transulphation

pathway is overwhelmed and folate trapped, homocysteine can be transported to the liver where the enzyme betaine methyltransferase (BMT); also known as betaine-homocysteine S-methyltransferase (BHMT; EC 2.1.1.5) is able to recycle it back to methionine (Steenge *et al.*, 2003). The reaction is dependent on betaine as the methyl donor and provided the diet is normal, the supply of betaine and/or its precursor, choline, is rarely considered. Khohlmeier *et al.,* (2005) however, raise the problem of low choline and betaine intake influencing folate metabolism and therefore must be considered as a potential influence on MTX resistance (Heil *et al.*, 2000; Meyer *et al.*, 2004).

1.14 AIMS

The central role of folate in normal cell metabolism and the complexity of its acquisition and use by the cells is now apparent. Overt intracellular deficiency of THF is the aim of antifolate therapy in ALL and other cancers, but the ability of cells to adapt and develop resistance is an on-going problem for a small but significant percentage of children with poor outcome. To improve this outcome there is a requirement for more detailed knowledge on the mechanisms of MTX resistance. The Paediatric Oncology Group at the Northern Institute of Cancer Research, Newcastle University, Newcastle-upon-Tyne, have an established research record that has enhanced the understanding of drug resistance in childhood ALL. Although it is well recognised that changes in expression of DHFR and RFC are common causes of MTX resistance, and a number of *in vitro* mutations had been studied, there is little information on *in vivo* mutations in these two proteins. The hypothesis being tested was that allelic imbalance and somatic mutations in genes involved in folate metabolism contribute to methotrexate resistance and subsequent relapse in childhood ALL.

Therefore the first aim of the work in this thesis was to:-

- Use denaturing high performance liquid chromatography to screen gDNA obtained from children with acute lymphoblastic leukaemia for somatic mutations in the reduced folate carrier and dihydrofolate reductase genes as potential causes of relapse. In order to meet this aim the following objectives were set:-
    a) Design suitable primers to amplify the coding exons of RFC and DHFR.
    b) Optimise PCR to obtain sufficient product for DHPLC analysis.
    c) Optimise DHPLC conditions.
    d) Interpret chromatograms and discuss findings.

The second aim was to:-

- Use gDNA Affymetrix SNP microarray data obtained from presentation and relapse childhood acute lymphoblastic leukaemia to identify potential allelic imbalance in genes involved in folate metabolism and possible MTX resistance. To meet this aim the following objectives were set:-
    a) Identify genes linked directly or associated with folate metabolism.
    b) Present data to show differences in estimated copy number between presentation and relapse samples.
    c) Present data to show difference in probability scores for loss of hetrozygosity between presentation and relapse samples.
    d) Discuss findings in difference between presentation and relapse samples for each gene, relationship between copy number and loss of heterozygosity; and possible contribution to methotrexate resistance and relapse.

2 SCREENING THE REDUCED FOLATE CARRIER

One of the aims of this research was to screen for mutations in the reduced folate carrier gene (RFC-1) from children who have relapsed and will be discussed in this chapter. Use of samples from this point in the disease may indicate if relapse is due to an acquired mutation developed during initial treatment and which now confers a degree of methotrexate resistance. To meet this aim the following objectives were established:-

a) Set-up and optimise PCR amplification of coding exons in the RFC-1 gene.

b) Optimise the denaturing high performance liquid chromatography screening method for each of the exons.

c) Amplify and screen a panel of gDNA available from children with ALL collected at relapse (sample most likely to have MTX related mutations) and compare this with gDNA from a panel of normal individuals

2.1 SCREENING FOR MUTATIONS AND POLYMORPHISMS

On the 14th April 2003, the International Human Genome Consortium announced the completion of the draft human genome sequence (National Human Genome Research Institute, 2003). However, the number of genes identified, or even proposed to lie, within the 2.9 billion bases do not account for the phenotypic diversity of the human population. The biodiversity that we see around us is the result of minor alterations in the genome, ie., single nucleotide polymorphisms (SNPs) (Cargill *et al.,* 1999). These often go unnoticed because they lie in the larger non-coding regions or are silent mutations; but some are significant enough to produce an evolutionary advantage or disadvantage. For this reason the scientific world has not sat back admiring its

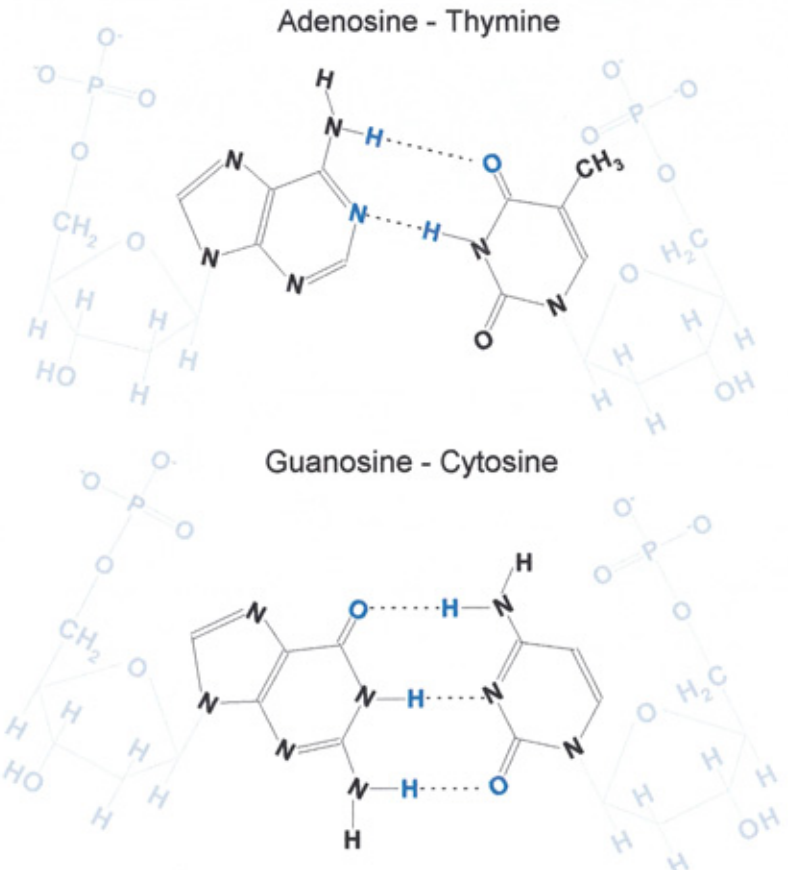
success, but has used the genome as the starting point to determine what the slight variations are and how these relate to disease.

The ultimate method for detecting and identifying mutations is to perform full sequencing (Kwok and Chen, 2003), but even with the advanced automated sequencing systems currently available, this is still unrealistic for large populations. In the last fifteen years, a range of methods have been developed each with faster throughput, sometimes at the expense of fidelity, but each have provided additional information to help our understanding of the genome (Cotton, 1997; Tabone *et al.,* 2006). Restriction mapping, using chemical or more specific restriction endonuclease activity, was one of the earliest methods and although it is really only useful to identify the presence of a known difference, it is still used for rapid population screening to determine the incidence of particular alleles. The last 5 years has seen a major shift in screening toward the use of DNA 'chip' microarrays that will be described in further detail in chapter 4. As the cost of microarrays continue to fall then this method is becoming more and more popular, but in the early days of screening it was the understanding of DNA's properties under specified conditions that saw a range of techniques developed. One of these, denaturing high performance liquid chromatography (DHPLC) fulfils many of the ideal criteria for a screening method (O'Donovan and McGuffin, 1997; O'Donovan, 1998). Since its introduction by Oefner and Underhill in 1995, the technique rapidly become the tool of choice in the rapid screening for previously unknown mutations and SNPs (Cotton, 1997; Underhill *et al.,* 1997; Liu *et al.,* 1998; Oefner and Underhill, 1998; Wagner *et al.,* 1999; Schriml *et al.,* 2000; Xiao and Oefner, 2001; Huber *et al.,* 2001; Frueh and Noyer-Weidner, 2003; Chou *et al.,* 2005) and will be used here to screen the RFC and DHFR genes.

2.2 DENATURATION OF DNA AND DUPLEX FORMATION

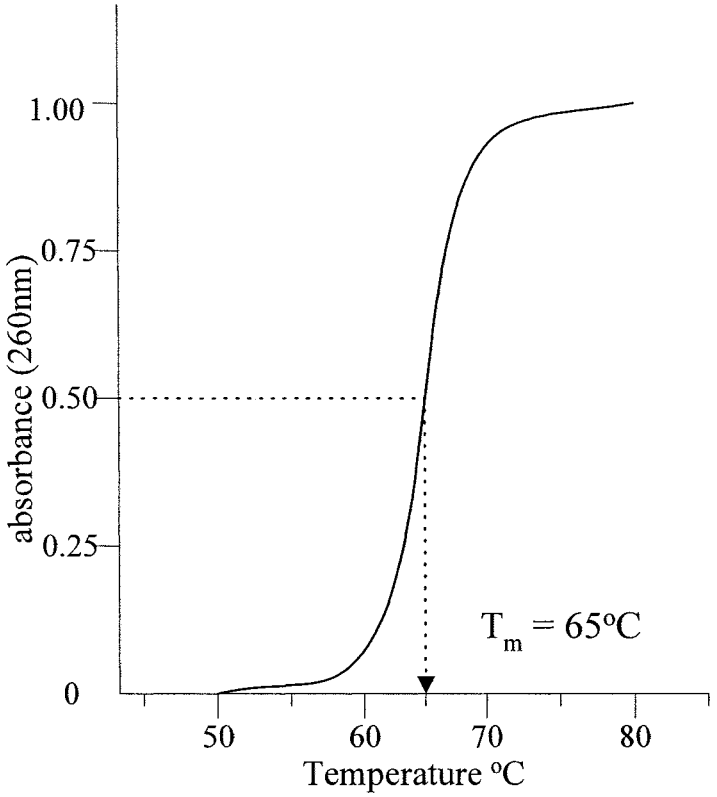
Destabilisation of the classic double helix (also known as denaturation or melting) is required in the majority of techniques used to study the genome and can be achieved relatively easily by heating the DNA. The thermal energy breaks the hydrogen bonding between paired bases, and as the double stranded duplex (dsDNA) is converted to single stranded DNA (ssDNA) this can be followed at 260nm (Marmur and Lane, 1960). The temperature at which a specific sequence of dsDNA is denaturated is dependent on its nucleic acid make up. G-C pairing with 3-hydrogen bonds is more stable than the two hydrogen bonds of an A-T pair (Figure 2.1) and thus a high percentage G-C content in a sequence produces a higher melting temperature.

Figure 2.1. Base pairing of adenosine with thymine and guanosine with cytosine (Watson and Crick; 1953a, 1953b). Hydrogen bonds are shown in blue, common sugar-phosphate moiety shown in grey.



When absorbance is plotted against temperature the resulting melting curve (Figure 2.2) shows a characteristic sigmoidal shape. This shape suggests that denaturation of one base makes the surrounding bases less stable and the rate of denaturation accelerates. The temperature at which half the change in absorbance is achieved is called the melting temperature (T_m).

Figure 2.2. Typical melting curve of DNA showing T_m.



The interbase hydrogen bonding by itself is too weak to explain the stability of the DNA duplex, but the bases forming the nucleic acid ring stacks along the helix have multiple weak van der Waals interactions and those with hydrophobic groups will attract and protect each other giving further strength. There are 10 possible dimer-stacking permutations (Table 2.1), with C-G/G-C requiring the most energy to dissociate and A-T/T-A the least (Melchior and von Hippel, 1973; Ke and Wartell,

1993). Other factors that influence the T_m are pH, ionic concentration and the presence of base pair mismatches (Marmur and Lane, 1960).

Table 2.1. Base stacking energies (kJ/mol) of all 10 possible dimer permutations (Taken from Ornstein *et al.,* 1978).

C-G G-C	-61.0
C-G A-T	-44.0
C-G T-A	-41.0
G-C C-G	-40.5
G-C G-C	-34.6
G-C A-T	-28.4
T-A A-T	-27.5
G-C T-A	-27.5
A-T A-T	-22.5
A-T T-A	-16.0

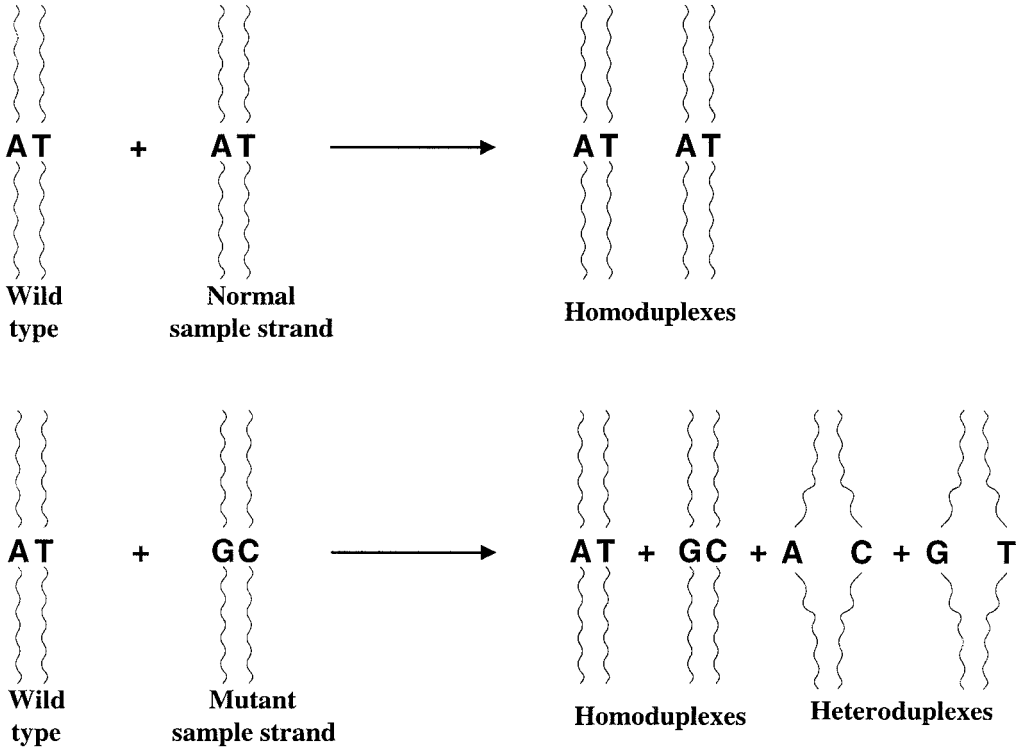
If denatured DNA is allowed to cool slowly, then the reannealing process ensures that matched (homologous) strands will reform perfectly into homoduplexes. However, if the sample contains a SNP/mutation in one of the alleles, then heteroduplexes (hybrids) will be formed as shown in figure 2.3.

Base pair mismatching due to substitutions, insertions and deletions alters the base stacking energies so that the T_m of the mutated DNA is different from the wild type and this can be differentiated by performing electrophoresis (Ke and Wartell, 1993) or

chromatographic separation (Oefner and Underhill, 1995) at a partial melting temperature. The temperature is critical and must be maintained within tight limits if reproducibility is to be obtained.

A SNP in its non-wild type homozygous form will also produce a homoduplex and a single peak. To check for this, the PCR amplification product is spiked with an equal amount of wild-type DNA and a cycle of denaturation and slow annealing performed (Marmur and Lane, 1960; Doty *et al.,* 1960). If the two are the same, then homoduplexes will again reform, if not, then heteroduplexes will be produced (Schildkraut *et al.,* 1961).

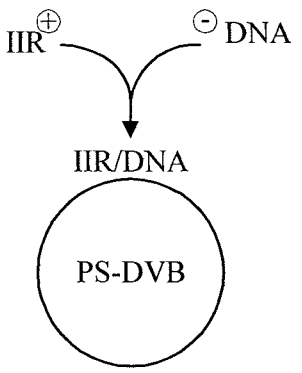
Figure 2.3 Diagram showing homoduplex formation when the sample strand of interest is the same as the wild type (top) and formation of four species, if the sample strand contains a SNP/mutation (bottom), two homoduplexes with matching bases and two unmatched heteroduplexes. (taken from WAVE® handbook; Transgenomic Inc).



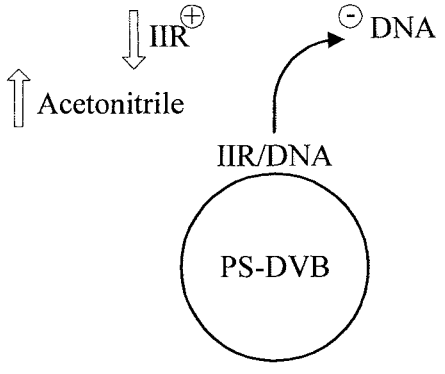
2.3 DENATURATING HPLC (WAVE®)

High performance liquid chromatography (HPLC) has been used since the 1980's for separating single strand DNA of different sizes, but its use in the separation of duplex DNA was only achieved with the synthesis of a nonporous polystyrene-divinylbenzene bead as the stationary phase (Huber *et al.,* 1992, 1993a, 1993b, 1993c, 1995; Oefner and Bonn 1994; Oefner and Underhill, 1995). This reverse phase system (ie polar mobile phase and non-polar stationary phase) was soon adapted for the separation of dsDNA fragments that not only differed by size but also by a single nucleotide (Huber and Berti, 1996). The mobile phase contains an ion-interactive reagent (IIR) that pairs with the negatively charged DNA making it non-polar (Robards *et al.,* 1994). The neutralised DNA then associates with the non-polar stationary phase and is retained on the column (shown on the left in figure 2.4).

Figure 2.4. Diagram showing ion-pairing and association with non-polar stationary phase (left) and effect of acetonitrile gradient (right). Taken from Robards *et al.,* 1994.



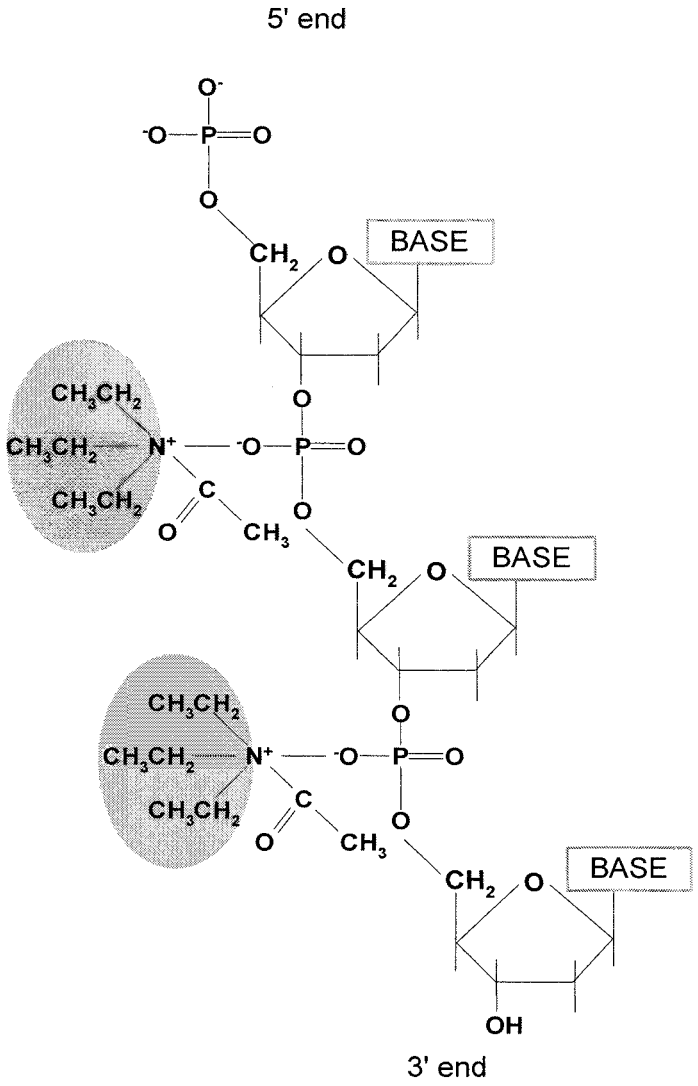
In the presence of the paired ion DNA is neutralised and will associate with the non-polar stationary phase.



As the acetonitrile gradient is applied and the amount of paired ion reduces, then DNA will be dissociated from the stationary phase and eluted.

Introducing a gradient that reduces the ion-pairing properties of the buffer disrupts ion-pairing and will cause the DNA to partition back into the mobile phase and its elution from the column. This is further enhanced by the introduction of an acetonitrile gradient into the mobile phase to modify its polarity. The rate of gradient change, start and end points can be selected for a specific DNA fragment (Oefner, 2000) and the eluted DNA detected using either UV at 254 nm (Oefner *et al*; 1994a) or fluorescent labels (Kosaki *et al.,* 2001; Bahrami *et al.,* 2002).

Figure 2.5. Diagram showing interaction of positively charged triethylammonium ions with the negatively charged phosphate backbone of DNA. The non-polar regions formed by the alkyl chains are shown as the shaded areas.



The positively charged ion (cation) chosen for the ion-pairing buffer needs to be able to fit into the large groove of DNA, where it binds to the negatively charged (anionic) phosphate (Jost *et al.,* 1982) (figure 2.5). Tetrapropyl and tetrabutyl-ammonium acetate buffers were found to be too large (Huber *et al.,* 1995) and of little value, whilst tetramethyl-, triethyl- and tetraethylammonium ions in TMA, TEA and TEAA buffers respectively, can. Any of these cation alkyl ammonium ions is suitable, however TEAA is used in this project. As the fragment size increases, more acetonitrile is required to elute it. Hence, retention time is also related to fraction size and gradient conditions (Huber *et al.,* 1995).

To find the optimum temperature, an empirical approach can be applied in which a heteroduplex is repeatedly injected at temperatures increasing by 3°C, starting at 50°C. At this low temperature both hetero- and homoduplexes will exist in the fully helical form and bind equally, eventually a temperature will be reached at which the heteroduplexes start to melt and binding to the stationary phase is reduced, giving it a shorter retention time. If the temperature continues to increase then remaining homoduplexes start to denature and will eventually be eluted (Giordano *et al.,* 1999). Altenatively, if the base sequence of the melting region is known an algorithm is available from the University of Stanford (*http://insertion.stanford.edu/melt.html*) or other sources (Jones *et al.,* 1999, Chou *et al.,* 2005) that use the stacking energies to predict the T_m for a given sequence. If there is more than one suspected melting domain, then the samples need to be run at this second temperature and where possible the use of a positive control is encouraged to ensure complete SNP detection (Giordano *et al.,* 1999).

2.4 WAVE® DHPLC SYSTEM

Transgenomic Inc (San Jose, CA, USA) developed a monolithic column, (DNASep®) which does not require frits to retain the original pelicular beaded stationary phase on the column. This reduces the backpressure on the system and allows freer mobile phase elution from the column and produces sharper elution peaks (Huber *et al.*, 2001). Whilst the column can be purchased individually, it is used in this study as part of an integrated HPLC system called the WAVE® Nucleic Acid Fragment Analyser System (Transgenomic Inc, San Jose, CA, USA). Xiao and Oefner (2001) showed that tight control of the temperature improves resolution, so the manufacturers have designed a system in which all critical components (column, buffers and injection valve are maintained within 0.1°C across a range of 35-80°C. An automatic sample holder gives fully automated injection capability of between 1 and 100 μL of sample from a 96 well microtitre plate.

The system is controlled from a personal computer via WAVEMAKER™ integrated software. This calculates the optimal T_m and gradient settings for analysis of a given dsDNA sequence, controls the pumps, oven and other instruments and records the detector output. In the improved WAVE®HS system, an integrated fluorimeter increases sensitivity and an enhanced software package, NAVIGATOR™, make it possible to perform multiplex analysis (Kosaki *et al.*, 2005) and manipulation of chromatograms to enhance identification of aberrant peaks (Colley *et al.*, 2005).

Optimisation of WAVE® conditions will be described for each exon of the reduced folate carrier later in this chapter and for DHFR in chapter 3.

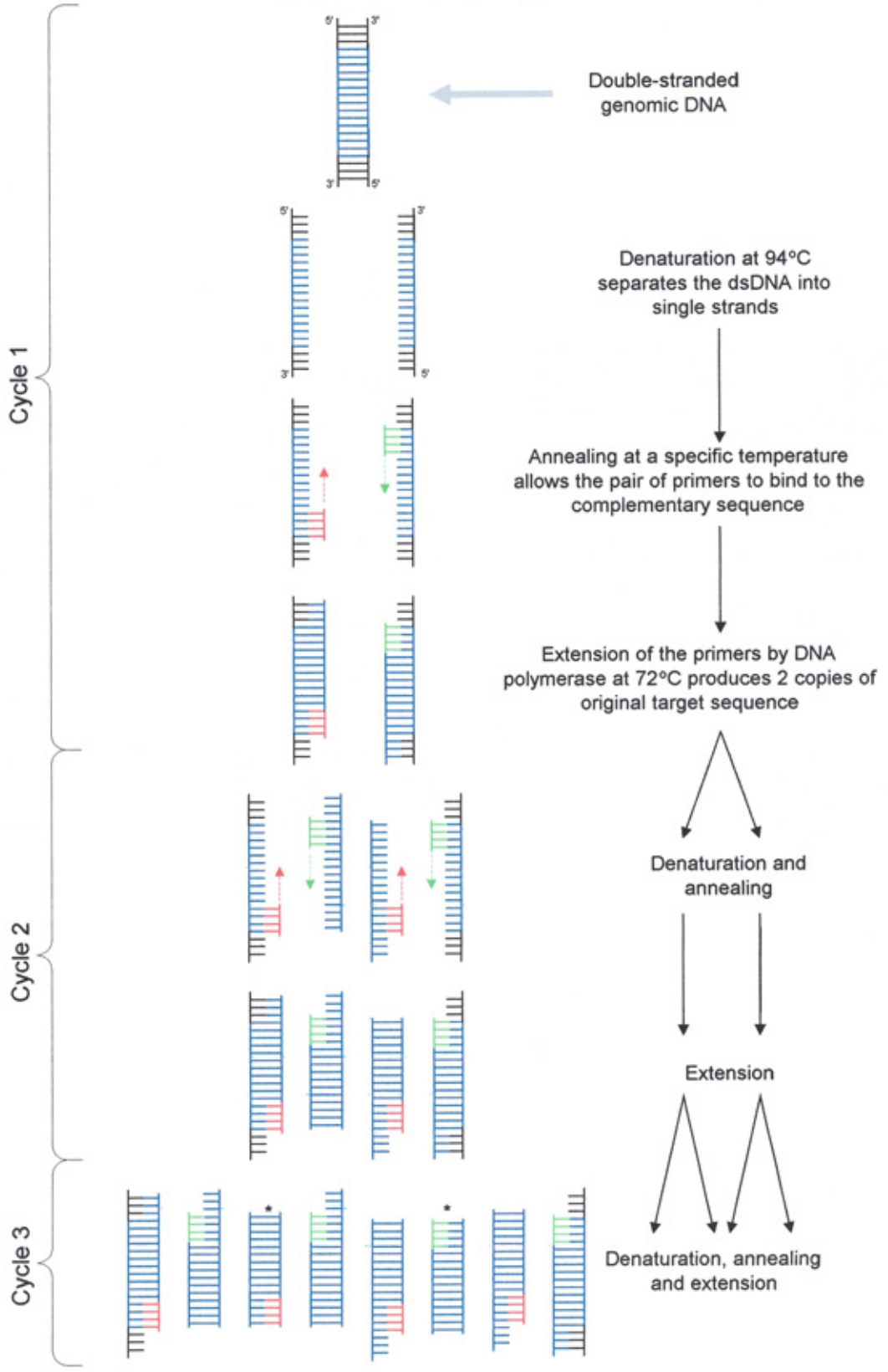
2.5 AMPLIFICATION OF THE RFC GENE

Whilst DHPLC is very sensitive (Oefner *et al.,* 1994b) it still requires selective amplification of the target dsDNA using PCR (polymerase chain reaction). Not only does this produce sufficient material for detection, but also produces a fragment of optimal size for separation, ie., 200-600 bp (Transgenomic Inc, San Jose, CA); any smaller and it is eluted too fast or denatures all at once (Huber and Berti, 1996), any larger and the resolution is compromised making interpretation more difficult. The following sections describe the design of primers for each of the RFC exons and optimisation of PCR conditions to produce a product suitable for WAVE® analysis.

2.5.1 POLYMERASE CHAIN REACTION (PCR)

PCR is a mainstay technique in molecular biology and uses the principle of logarithmic amplification of a specific sequence of dsDNA (Saiki *et al.,* 1985; Mullis and Faloona, 1987). The target sequence is defined by a pair of primers, one lying either side of the target sequence but on opposite strands (figure 2.6). In the first cycle of PCR the original dsDNA is denatured at 94°C into single strands. As the temperature is allowed to cool the primers anneal to complementary sequences on the DNA. If primers have been designed carefully this will be either side of the target sequence and no other points on the genome. The temperature at which this occurs is defined for each primer and for maximum efficiency the primers should anneal at the same temperature or within 1°C of each other. The cycle then progresses into the extension step in which DNA polymerase uses the primer on each strand as the origin and using deoxynucleotide triphosphates (dNTPs) available in the mixture, begins to build a complementary strand to each of the originals. On completion of the first cycle there are two copies of the target sequence.

Figure 2.6 Diagram showing denaturation, annealing and extension steps in polymerase chain reaction (PCR) through three cycles. The target sequence is shown in blue, forward primer (sense) is red and reverse primer (antisense) is green. Note that for diagrammatic purposes the sequence lengths of primers and target DNA are shorter than actual lengths.



This process is repeated 20-40 times and on each cycle the copy number is doubled. Also, because one of the copy strands is limited by the primer, the products become more specific for the target sequence (see products labelled * in figure 2.6) and eventually these 'unit length strands' become the dominant product amongst the millions of copies produced (Russel, 2002).

The annealing temperature is specific for each pair of primers; too high and the primers fail to anneal, too low and they bind non-specifically and generate several products. Although a proposed temperature can be calculated it is usually determined empirically. The second parameter that can often be optimised in PCR is the concentration of magnesium. This divalent cation is required as a cofactor to control which phosphate-phosphate bond is cleaved in the dNTP being added and ensures pyrophosphate is released.

2.5.2 PRIMER DESIGN

Initial literature searches in 2001 produced few papers with references to primers for use in amplifying human RFC genomic DNA (gDNA) (Gong *et al.,* 1997, 1999; Jansen *et al.,* 1998; Zhang *et al.,* 1998a,1998b; Chango *et al.,* 2000; Drori *et al.,* 2000a, 2000b). Primers proposed by Zang *et al* (1998a, 1998b), Gong *et al* (1997;1999) and Chango *et al* (2000) produce a product or products too big for WAVE® analysis. Of the remaining papers, they cited a chapter written by Jansen (1999) as the source of information regarding experimental conditions for PCR amplification. However the chapter gives no primers and as a result primers had to be designed from scratch in June 2001. Later that year Whetstine *et al* (2001b) published six pairs of primers covering the coding exon.

In the design of primers, sequences on the EMBL database were identified that contain intronic information up and downstream of the required exons; U92869 for exon 2, U92870, U92871, U92872 and U92873 for exons 3, 4, 5 and 6 respectively. However, in all cases the amount of intronic information either side of the exon proved to be insufficient to facilitate effective primer design. As a result, the whole 340 kb segment covering the q22.2-q22.3 region of chromosome 21 (HS21C103; accession number AL163303) was downloaded from the NCBI website (*http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=7717449*) and copied into Omiga 2.0 (Accelrys, San Diego, CA) and DNAStar (Madison, WI) sequence handling software packages. Sequences for exons 2-6 were also downloaded from the NCBI website, saved to these packages and each exon subsequently aligned to confirm the correct region of chromosome 21 was being used. Failure of the initial alignment analysis indicated that the large HS21C103 sequence required transformation to its reverse complementary sequence. Alignment of each exon against HS21C103 allowed the selection of an appropriate section that provided sufficient up and downstream intronic data to facilitate primer design (table 2.2).

Table 2.2 Shows the region of HS21C103 from chromosome 21 used in primer design for exons 2-6 of the RFC-1 gene and the NCBI accession number for each exon used in the alignment.

RFC exon number (using Tolner's model)	Region of HS21C103 used for alignment and design of primers	Exon accession number used for alignment
2	236001 - 237300	U92868
3	230173 - 230944	U92869
4	229554 - 229767	U92870
5	224604 - 224743	U92871
6	213483 - 214933	U92872

Any of the primers considered suitable for use were entered into the NCBI BLAST search facility (*http://www.ncbi.nlm.nih.gov/blast/*) to check that the primers did not align elsewhere on the genome and potentially cause non-specific product amplification.

2.5.2.1 **Design of exon 2 primers**

DNAStar was initially used for primer design, but all the primer pairs proposed would generate products >1000 bp in length and therefore were considered unsuitable for WAVE® analysis. Omiga 2.0 uses an alternative algorithm for primer design and allows modification of algorithm parameters to increase the chance of design, or specify particular criteria for the product (eg., generate a final sequence which matches optimum WAVE® conditions). The majority of primer pairs proposed also generated a product >1000 bp in length and therefore were rejected. Of the few remaining pairs none were considered perfect, but to move the project forward, a pair with a ΔT_m of 0.1°C and a product length of 691 bp was selected (preparation details are shown in appendix 1). These primers subsequently failed to generate PCR products in several optimisation experiments and were rejected in 2002 and hence are not shown here. By this time the Whetstine *et al* (2001b) paper had been published and their primers were used for further work on exon 2.

RFC Ex2 forward	CTG ACT CCA CCC CTC CTT CCA GGC ACA GTG
RFC Ex2 reverse	CCA CAT GCC TGC TCC CGC GTG AAC TTC T

These primers generated a product 268 bp in length shown in Appendix 2.

2.5.2.2 **Design of exon 3 primers**

Exon 3 is a relatively large exon and as a single PCR product was considered too big for effective analysis on the WAVE®. Amplification of two smaller overlapping products was thought the most effective option and two sets of primers were designed. The following primer pairs designed by DNAStar were considered the most suitable for further use (Ex3.1 covers the 5'-end as shown in Appendix 3 and Ex3b the 3'-end).

RFC Ex3.1 forward	GTT CAG AGA GGA CGA TGC AG
RFC Ex3.1 reverse	GAA GGT GAG GAA GGC CAG

These primers produced a 569 bp product and gave a strong band of the expected size and following purification and sequencing was confirmed to be the 5' end shown in Appendix 3.

RFC Ex3b forward	CGG TCG TCG CGG TTG AAG AAG A
RFC Ex3b reverse	ACC TCG GGC GGG CTG TGT C

These primers should have produced a 536 bp product.

The Ex3b primers proved difficult to optimise and did not generate a product of sufficient intensity. As a result, the Whetstine primers (Whetstine *et al.*, 2001b) were considered, but on closer scrutiny did not quite cover the whole of the coding region (starts at ^{194}CGA^{196} and misses the first 4 bases GTCA that make up a critical splice-site) and therefore were not considered appropriate for this project.

With the poor success of the DNAStar primers and the unsuitability of the Whetstine primers a second attempt at designing primers using Omiga was made using wider design parameters and the following primer pairs were suggested.

RFC Ex 3.2 forward	AGC TCT TCT ACA GCG TCA CCA T
RFC Ex 3.2 reverse	TCT GGG GGA AGG TAT CCA T

This pair should have generated a 709 bp product of the 5' end of exon 3, but again, did not facilitate amplification of a suitable product. In 2004, after several attempts using other software a successful set of primers were designed using freeware from Technische Universitat Munchen (*http://ohg.gsf.de/cgi-bin/primer/ExonPrimer.pl*).

RFC Ex t3.2 forward	TCT CCT TCT CCA CGC TCA AC
RFC Ex t3.2 reverse	AAG AAG CCT CGG GGA CC

These primers generated a 441 bp product covering the 5' end of exon 3 shown in appendix 3

2.5.2.3 **Design of exon 4 primers**

Neither DNAStar or Omiga software packages were able to suggest suitable primer pairs even after widening the design criteria and therefore those proposed by Whetstine *et al* (2001b) were ordered and used.

RFC Ex4 forward	CCC CTC TCA GGC GCC ATC ACG TCC TTC
RFC Ex4 reverse	GTG GAG GGC CTG GGG GAG CAG CAA

These primers generated a product of 427 bp and are marked on the sequence shown in appendix 4.

2.5.2.4 **Design of exon 5 primers**

A pair of primers giving a product of 250 bp were generated on the first attempt using Omiga. Primers are marked on the sequence for exon 5 shown in appendix 5.

RFC Ex5 forward	CTT ATT CTC CCA GCA GCG CC
RFC Ex5 reverse	GCC TCA ACA ATG TCC CCA CA

2.5.2.5 **Design of exon 6 primers**

The Whetstine *et al* (2001b) primers for this exon were considered, but when aligned, the forward primer starts 12 bases into the 5' end of the exon. As this could potentially miss any SNP or mutation along this small stretch it was decided not to use these primers and try to design a new pair using Omiga software.

RFC Ex6 forward	CTG AGC TGG AGG AGG GTG A
RFC Ex6 reverse	CCC CAT TGC TAA GGC AGG

Despite several experiments to optimise conditions, amplification using these primers proved unsuccessful. DNAStar was unable to generate suitable primers, but the PRIMO PRO3.4 shareware package available from Chang Biosciences (*http://www.changobioscience.com/primo/primo.html*) generated several possibilities with a product size at the maximum considered suitable for WAVE® analysis. From these the following pair giving a fragment of 649 bp was selected and ordered for further work (shown in appendix 6).

RFC p6 forward	TGA CCG CTG TGC AGT GTT CAG
RFC p6 reverse	CGC CTG CAA AGT TAC CAC AGG

Primers were obtained from MWG Biotech AG (Covent Garden, England).

2.5.3 PREPARATION OF PRIMERS

All primers were received as lyophilised vials and stored at 4°C until reconstituted with 1 mL of sterile 18.2 MΩ deionised water. When fully dissolved, 5 x 100 µL aliquots were prepared and stored at –20°C as stock vials, whilst the remainder of the reconstituted primer was stored at –80°C. From the stock primer solution, 5 µmol/L working solutions were prepared and used immediately or stored at –20°C (see appendix 1 A4).

2.5.4 OPTIMISATION OF PCR CONDITIONS

To obtain the maximum efficiency from a DNA polymerase in PCR it is necessary to optimise the amplification temperature and magnesium concentration. The main DNA polymerase used in this project was AmpliTaq Gold® (Applied Biosystems, Foster City, CA) at an activity concentration of 5 U/µL. The Tris-HCl buffer is supplied ready for use as a 10x concentrate (100 mmol/L Tris-HCl, pH 8.3, 500 mmol/L KCl). Magnesium is supplied as a separate 25 mmol/L $MgCl_2$ solution to allow titration of the optimum concentration. To prevent misincorporation of nucleotides the four deoxyribonucleotide triphosphates (dATP, dGT, dCTP and dTTP) were added as an equimolar master mix (100 mmol/L dNTP Master Mix; Bioline, London, UK), diluted and used as a 12.5 mmol/L working solution. All water used in PCR was sterilised 18.2 MΩ deionised water and all consumables were either autoclaved in-house or purchased as DNA-free.

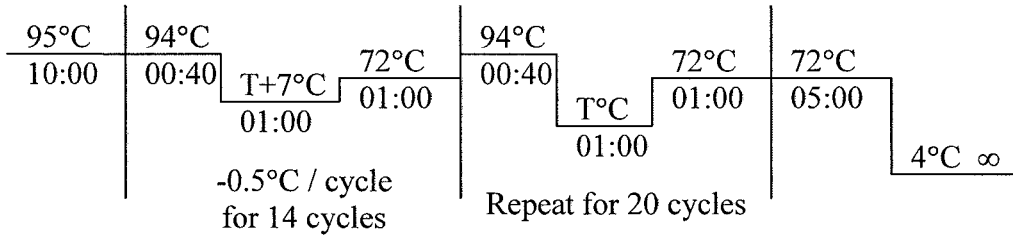
Initial PCR for all primer pairs used the manufacturer's recommended reagent concentrations and thermal cycler conditions for amplification using AmpliTaq Gold® (Roche Molecular Systems, 2000) as shown in table 2.3 and figure 2.7. These were subsequently modified for each exon as described in section 2.5.4.2 – 2.5.4.6.

AmpliTaq Gold® is a high fidelity thermal stable DNA polymerase which has been modified by the addition of an inhibitor to inactivate it at low temperature. This prevents non-specific amplification as the temperature is ramped up to the denaturation temperature. The initial 10 minute incubation of the PCR mixture at 95°C is to uncouple the inhibitor and activate the polymerase which is most efficient at 72°C. To improve specificity, 'touch-down' PCR (Don *et al.*, 1991) was applied unless otherwise stated.

Table 2.3 Working concentrations and volumes of reagents used in the initial PCR run for new primer pairs. Total volume = 50 µL.

	Original concentration	Working concentration	Volume (µL)
Buffer	10x	1x	5.0
$MgCl_2$	25 mmol/L	2.5 mmol/L	5.0
dNTP Master Mix	12.5 mmol/L	100 µmol/L	0.4
Forward primer	5 µmol/L	0.2 µmol/L	2.0
Reverse primer	5 µmol/L	0.2 µmol/L	2.0
DNA	Approximately 50 ng		1.0
AmpliTaq Gold®	5 U/µL	1.25 U/50µL	0.25
Water			34.35

Figure 2.7 Initial PCR conditions, (where T is the proposed primer annealing temperature and times are shown below the line in minutes:seconds).



2.5.4.1 **Optimisation of annealing temperature**

As already mentioned, the optimum annealing temperature is often determined empirically despite a proposed temperature being calculated by the primer design software. The Eppendorf Mastercycler® (Cambridge, UK) is capable of running a temperature gradient across the heated plate and allowed temperature optimisation to be achieved in a single run. This technique was applied for each pair of primers as described in sections 2.5.4.2 – 2.5.4.6. Success of PCR was assessed using agarose gel electrophoresis and ethidium bromide visualisation as described in section 2.5.5.

2.5.4.2 **Optimisation of RFC exon 2 amplification**

A temperature gradient ranging from 57 – 67.5°C showed that 59°C produced the greatest amount of product, but at this low temperature was also starting to generate non-specific products. Specificity was improved by using a 'touchdown' modification of the PCR run (Don *et al.*, 1991) (see figure 2.8). In this modification, higher annealing temperatures are used in the early cycles so that primers anneal with lower efficiency but greater stringency. The temperature was then reduced at the annealing step by 0.5 or 1.0°C over a number of cycles until the temperature for optimum replication efficiency was reached. The early cycles increase the copy number of the specific target sequence so that when the lower target temperature is reached there is proportionally more of the target being amplified at each cycle compared to non-specific sequences.

To further refine amplification a magnesium gradient was set up covering the range 0.5 – 5.0 mmol/L and a final concentration of 2.0 mmol/L $MgCl_2$ was found to be the optimum for this exon. Final volumes and conditions used in the PCR are shown in table 2.4 and figure 2.9

Figure 2.8 Agarose gels showing standard PCR (left) and 'touchdown' PCR (right) at various temperatures; Lane 1 = 61.1 °C, lanes 2 and 5 at 58.6°C, lanes 3 and 6 = 57.0°C, lane 4 = negative control, to show inproved specificity of PCR when using 'touchdown' protocol.

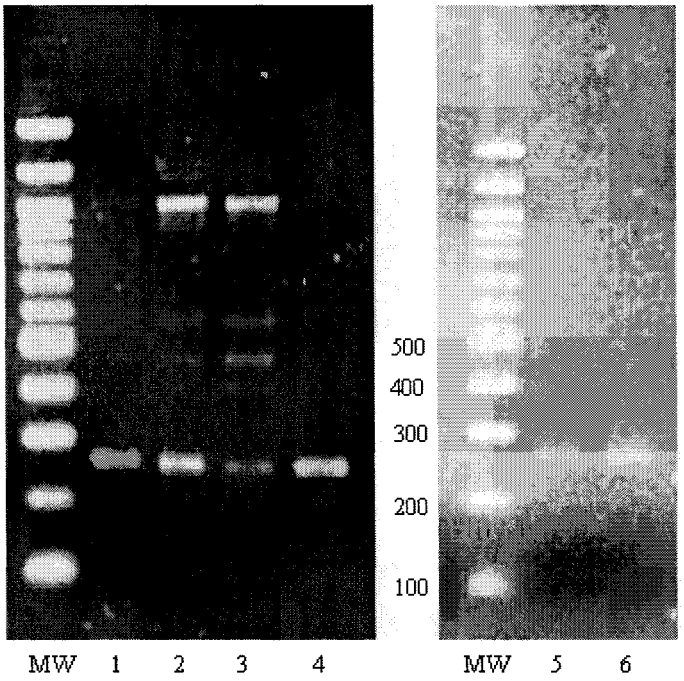
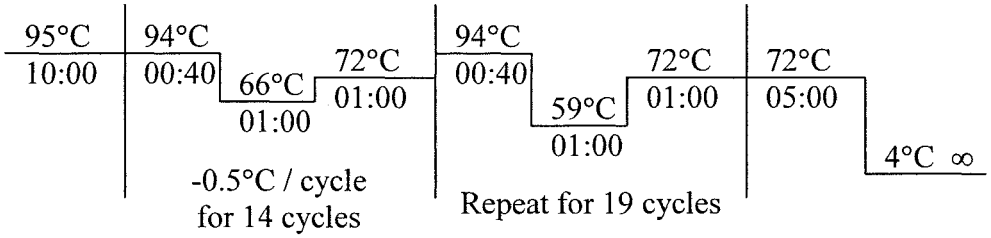


Table 2.4 Volumes and concentrations of reagents used in PCR amplification of RFC exon 2 . Total volume = 50 µL.

	Original concentration	Working concentration	Volume (µL)
Buffer	10x	1x	5.0
$MgCl_2$	25 mmol/L	2.0 mmol/L	4.0
dNTP Master Mix	12.5 mmol/L	100 µmol/L	0.4
RFC Ex2 forward	5 µmol/L	0.2 µmol/L	2.0
RFC Ex2 reverse	5 µmol/L	0.2 µmol/L	2.0
DNA	Approximately 50 ng		1.0
AmpliTaq Gold®	5 U/µL	1.25 U/50µL	0.25
Water			35.35

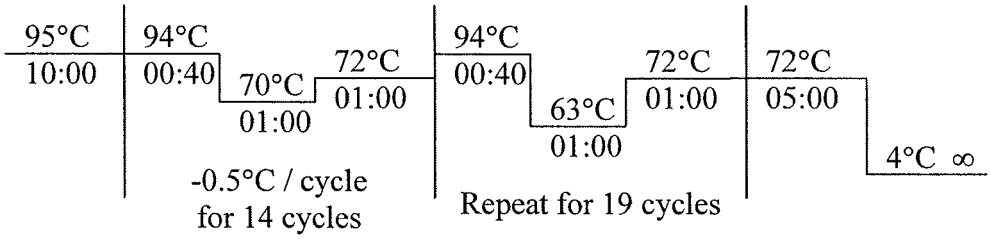
Figure 2.9 PCR conditions for RFC exon 2 amplification. (Times in min:secs)



2.5.4.3 **Optimisation of RFC exon 3 amplification**

Optimising amplification conditions of the 5' end using the RFC Ex 3.1 primer pair was relatively uneventful. Concentration and volumes of PCR reagents are those shown in the standard preparation (table 2.3), temperatures and timings used during amplification are shown in figure 2.10.

Figure 2.10 PCR conditions for RFC exon 3.1 amplification. (Times in min:secs)



The 3' end of exon 3 proved more difficult to amplify. Although the RFC Ex t3.2 set of primers gave a product of the expected size, optimisation of PCR conditions by performing a temperature gradient, varying magnesium, primer and DNA concentrations had little effect on increasing amplification efficiency. The addition of betaine as a PCR enhancer produced an immediate benefit (figure 2.11) and was adopted into the PCR mixture as shown in table 2.5. PCR timings and temperatures are shown in figure 2.12.

Figure 2.11 Agarose gel showing effect of betaine addition to PCR mixture in amplification of RFC exon t3.2. (Lane 1 = negative control, lane 2 = 1% $^{w}/_{v}$ betaine, 3 = 2%, 4 = 3%, 5 = 4%. 6 = no betaine added; 400 and 500 base pair bands marked on molecular weight ladder).

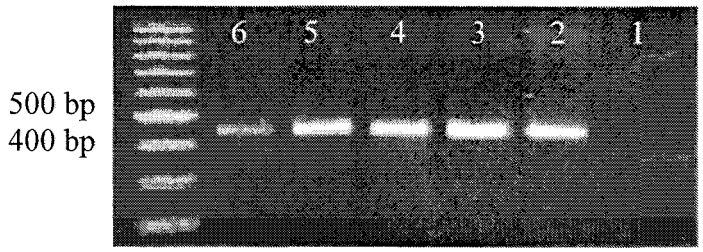
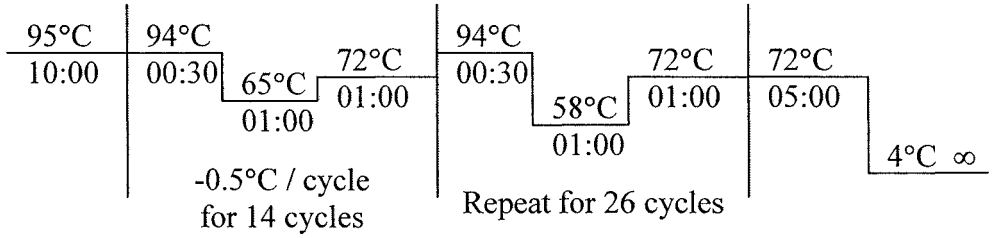


Table 2.5 Volumes and concentrations of reagents used in PCR amplification of RFC exon t3.2 . Total volume = 50 µL.

	Original concentration	Working concentration	Volume (µL)
Buffer	10x	1x	5.0
$MgCl_2$	25 mmol/L	2.0 mmol/L	4.0
dNTP Master Mix	12.5 mmol/L	100 µmol/L	0.4
RFC Ex2 forward	5 µmol/L	0.2 µmol/L	2.0
RFC Ex2 reverse	5 µmol/L	0.2 µmol/L	2.0
DNA	Approximately 100 ng		1.0
AmpliTaq Gold®	5 U/µL	1.25 U/50µL	0.25
Water			25.35
Betaine	10 % $^{w}/_{v}$	2 % $^{w}/_{v}$	10.0

Figure 2.12 PCR conditions for RFC exon t3.2 amplification. (Times in min:secs)



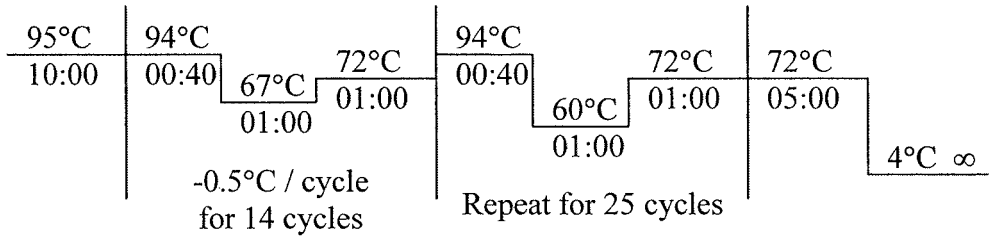
2.5.4.4 **Optimisation of RFC exon 4 amplification**

The exon 4 primer pair gave a product of the appropriate size but optimisation was proving very difficult. The addition of betaine had a significant effect on amplification efficiency and provided the reproducibility missing in earlier experiments and now allowed optimisation of PCR conditions defined in table 2.6 and figure 2.13.

Table 2.6 Volumes and concentrations of reagents used in PCR amplification of RFC exon 4 . Total volume = 50 µL.

	Original concentration	Working concentration	Volume (µL)
Buffer	10x	1x	5.0
$MgCl_2$	25 mmol/L	2.5 mmol/L	5.0
dNTP Master Mix	12.5 mmol/L	100 µmol/L	0.4
RFC Ex2 forward	5 µmol/L	0.2 µmol/L	2.0
RFC Ex2 reverse	5 µmol/L	0.2 µmol/L	2.0
DNA	Approximately 100 ng		2.0
AmpliTaq Gold®	5 U/µL	1.25 U/50µL	0.25
Water			8.35
Betaine	10 % $^{w}/_{v}$	5 % $^{w}/_{v}$	25.0

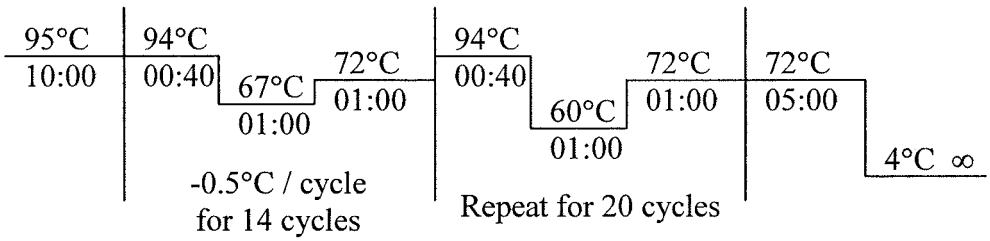
Figure 2.13 PCR conditions for RFC exon 4 amplification. (Times in minutes)



2.5.4.5 **Optimisation of RFC exon 5 amplification**

Concentration and volumes of reagents used in the PCR mixture were those shown in the standard preparation (table 2.3), temperatures and timings are shown in figure 2.14.

Figure 2.14 PCR conditions for RFC exon 5 amplification. (Times in minutes)



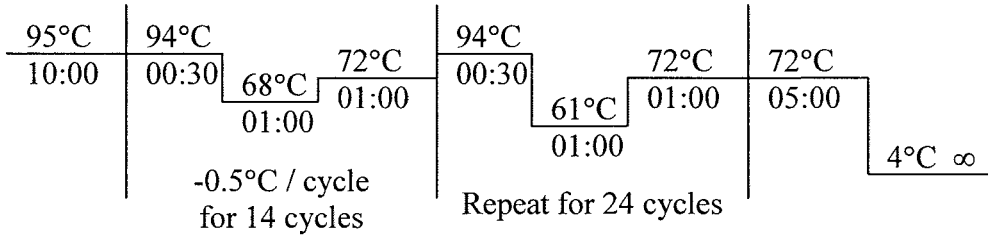
2.5.4.6 **Optimisation of RFC exon 6 amplification**

Amplification of this exon proved the most difficult. Over a period of almost a year a series of experiments to optimise conditions failed, even with the addition of betaine and/or DMSO as PCR enhancers. In 2004, Transgenomic introduced a new high fidelity DNA polymerase, called Optimase™ (product number 703050; Transgenomic Inc, San Jose, CA) which was reported to be successful in amplifying GC rich sequences and had been approved for use on the WAVE® (Personal communication and later Papadopoulou *et al.,* 2004). The first use of this polymerase, using the manufacturer's specified conditions, gave a much improved amplification and did not require any further optimisation. The concentrations and volumes of reagents are shown in table 2.7, (note that the buffer contains 15 mmol/L $MgSO_4$); PCR conditions are shown in figure 2.15.

Table 2.7 Volumes and concentrations of reagents used in PCR amplification of RFC exon 6 . Total volume = 50 µL.

	Original concentration	Working concentration	Volume (µL)
Buffer	10x	1x	5.0
dNTP Master Mix	12.5 mmol/L	0.8 mmol/L	0.4
RFC Ex p6 forward	5 µmol/L	0.8 µmol/L	8.0
RFC Ex p6 reverse	5 µmol/L	0.8 µmol/L	8.0
DMSO	100 % $^{V}/_{V}$	2.5 % $^{V}/_{V}$	1.26
DNA	Approximately 100 ng		2.0
Optimase™	2.5 U/µL	2.5 U/50 µL	1.0
Water			23.6

Figure 2.15 PCR conditions for RFC exon 6 amplification. (Times in minutes)



2.5.5 SEPARATION AND VISUALISATION OF PCR PRODUCTS

The outcome of each optimisation experiment was determined by separation of any PCR products by electrophoresis on a 2% agarose gel at 100 volts until dye markers showed sufficient migration (usually 30 – 45 minutes). Agarose gels were prepared by dissolving 1 g of powdered molecular biology grade agarose in 50 mL 1x Tris-borate EDTA (TBE) buffer by heating the solution in a microwave; 25 µL ethidium bromide (1 mg/mL) was added and the mixture poured, whilst still molten, into a casting tray fitted with an appropriate sized coomb. Once set, the gel was placed into an electrophoresis tank, covered with 1x TBE buffer and the comb removed. To assist delivery of the sample into the well and to visualise progression of electrophoresis with the naked eye, 10 µL of PCR product was mixed with 2 µL of loading dye (G1881 6x loading dye; Promega, Southampton, UK). Following electrophoresis, the DNA in the gel was visualised under UV light (LePecq and Paoletti, 1967) and the image captured electronically on the BioRad Gel Doc 2000 system (BioRad Laboratories Ltd, Hemel Hempstead, UK) or using Poloroid film (Polaroid Ltd, Luton, UK).

At least one molecular weight marker is run alongside the samples to facilitate recognition of appropriate bands in the test samples. Markers used in this work were either;

- G3161 (Promega, Southampton), for use, mix 5 µL with 1 µL G1881 loading dye;
- P9577 (Sigma, Poole, UK), for use, mix 5 µL with 1 µL supplied loading dye;
- 100 bp ladder (Product number 15628-019; Invitrogen Ltd, Paisley), supplied ready for use.

By measuring the distances each band travels and plotting the log distance against size, a standard curve was produced from which a more accurate estimate of unknown product size was made. This was done for the product generated by each set of primers at an early stage of optimisation to determine if the product being generated was of the expected size. Amplification of the correct sequence was later confirmed by direct sequencing of the product.

2.5.6 CONFIRMATION OF SEQUENCE

Once PCR conditions had been optimised to give a sufficiently intense band the product obtained from amplification of NALM6 gDNA was purified using the QIAquick PCR Purification Kit according to the manufacturer's protocol (Product code 28104; QIAGEN, Crawley, UK). 500 μL of buffer PB was added to 100 μL of PCR product and vortex mixed. The mixture was then transferred to a QIAQuick column held in a capture tube and spun in the Joan A14 microcentrifuge at 13,000 rpm for 1 minute. The flow-through in the capture tube was discarded and 750 μL of buffer PE containing ethanol added to the spin column. Following centrifugation at 13,000 rpm for another minute the QIAQuick column was transferred to a sterile 2 mL Eppendorf tube. The final elution/concentration step was performed by adding 30 μL buffer EB carefully to the column filter and centrifuging for 1 minute at 13,000 rpm. The purified product was sent with accompanying 60 pmol/L solutions of each primer for sequence analysis on the ABI Prism 377 DNA Sequencer (Newcastle Medical School, Newcastle-upon-Tyne, UK).

In this procedure each primer was used in turn to amplify the matching product in the presence of fluorescently labelled ddNTP and normal dNTPs. At each base along the sequence a proportion were inhibited from further extension by the chance

incorporation of appropriate ddNTP. After several cycles a mixture of fluorescently labelled products each differing by one base pair was produced. These were then separated via gel electrophoresis and as the fluorescently labelled bands migrated from the gel they were read by laser and recorded. This was repeated for the reverse primer to get both the forward and reverse complementary sequences of NALM6 cell-line DNA. NALM6 was selected as the wild-type source of gDNA because it was readily available in the laboratory and the cell-line is not known to show MTX resistance in its wild-type state. The sequence report generated (example shown in appendix 2.2) was then loaded into Omiga and alignment performed. In each case, alignment confirmed that the target sequence containing the relevant exon had been amplified.

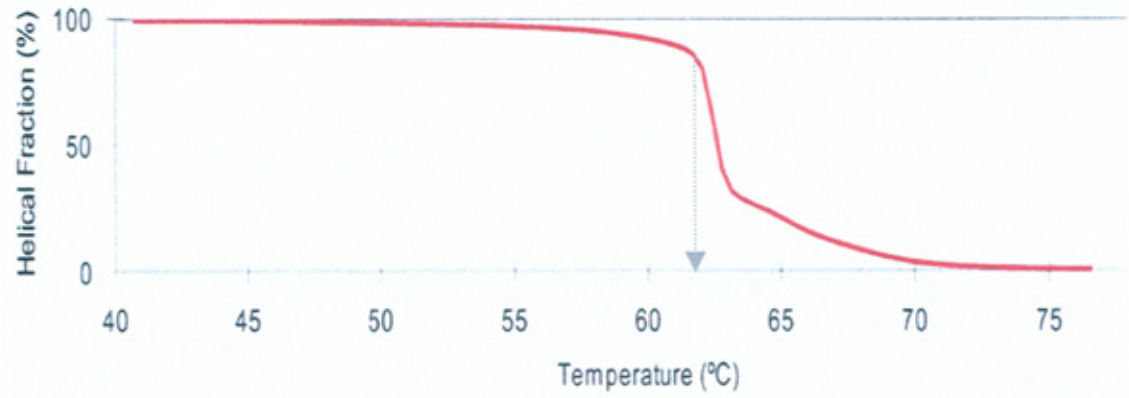
2.5.7 OPTIMISATION OF WAVE® CONDITIONS

Following confirmation that the amplified sequence was that of interest, the PCR product of each exon amplified from NALM6 was used to check DHPLC conditions. The expected wild-type sequence of the product was entered into WAVEMAKER™ or NAVIGATOR™ (after August 2004) and the software generated the optimum temperature. This selects the temperature at which 75% of the dsDNA is in its helical form (figure 2.16).

Based on the previous success within the research group of WAVEMAKER™ being able to predict the ideal partial melting temperature, early analysis used the predicted temperature. This approach was modified after problems with multiple melting domains within the sequences were identified and the advanced functions of the NAVIGATOR™ software were utilised to produce a detailed melt map (see figure

2.18). Using this information one or more temperatures were used in subsequent analysis and will be described in sections 2.5.7.2 – 2.5.7.6 for each exon.

Figure 2.16 Graph generated by WAVEMAKER™ showing % helical fraction against temperature for RFC exon 5 primer product. Predicted temperature at which 75% dsDNA exists in the helical form is shown by the arrow.



2.5.7.1 **Preparation of WAVE® system**

The expected sequence for the fragment to be analysed was entered into or retrieved from the method log on the hard-drive in WAVEMAKER™ or NAVIGATOR™ software and parameters for gradient and general run conditions were generated (see example shown in table 2.8). Routine pre-run maintenance was performed according to the manufacturer's instructions. Standards and test sample positions were programmed into the software and samples loaded. The remainder of the process is fully automated with each sample taking 8-9 minutes on the WAVE® or 2.5 – 3.0 minutes on the faster throughput WAVE®HS. Chromatograms were visualised and further analysed using WAVEMAKER™ or NAVIGATOR™ software as described in the following sections.

2.5.7.2 **Optimisation of WAVE® conditions for analysis of exon 2 PCR products**

The predicted temperature of 64°C was used for the initial analysis and the single peak obtained from NALM6 was that expected from a homozygous sequence (figure 2.17). However it is important to confirm that the conditions specified for the WAVE® can detect the presence of heteroduplexes by using a positive control (Schollen *et al.,* 2005). Following a search of the literature in September 2002, Dr Gerritt Jansen (Department of Oncology, University Hospital Vrije Universiteit, 1081 HV Amsterdam, The Netherlands) was contacted and he kindly donated a methotrexate resistant cell line. This CEM/MTX cell line was reported to contain a G to A substitution in one allele of exon 2 at position 133 (counting from the start codon) (Jansen *et al.,* 1998). DNA was extracted from a portion of the cells using the QIAamp DNA minicolumn extraction kit (product code 51104; QIAGEN Ltd, Crawley, UK), amplified and analysed. WAVE® analysis of this DNA clearly demonstrated the presence of a heteroduplex (chromatograms not shown) and this was accepted as a validation of the predicted temperature and amplification of patient samples went ahead.

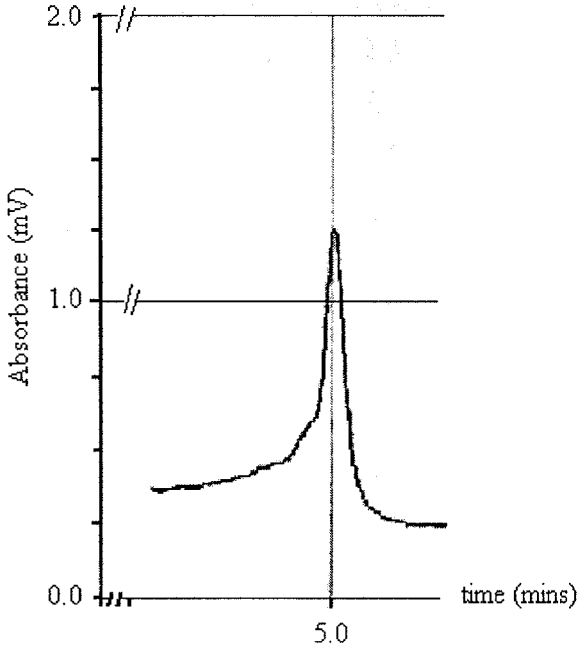
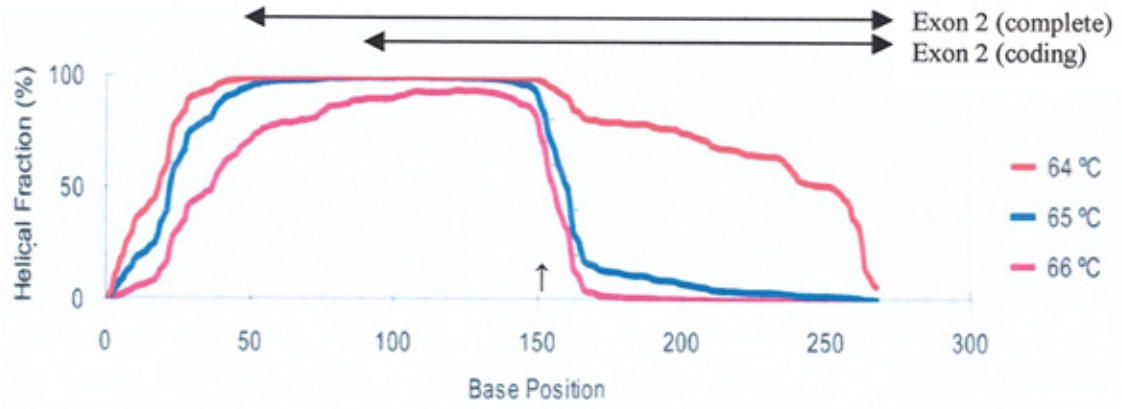


Figure 2.17 Homozygous peak obtained for RFC exon 2 at 64°C from NALM6.

However, following first run analysis of the normal and relapse samples at 64°C it was clear that the method had failed to pickup the recently identified common G80A polymorphism (Whetstine *et al.*, 2001b) and further laboratory work was performed to refine the optimum conditions. Employing the empirical approach (ie., PCR products were rerun at 64°C ± 3°C) the expected heteroduplex pattern expected from the G80A SNP was still not detected. Analysis of the WAVEMAKER™ generated melt map (figure 2.18 shows the melting temperature at a specific position calculated to take into account the influence of the stacking energies of surrounding bases presented graphically) shows that there are two distinct melting domains.

Figure 2.18 Melt map generated by WAVEMAKER™ software for the RFC exon 2 fragment showing the percentage of helical DNA at a given position at three temperatures (64, 65 and 66°C). Annotated arrows indicate the regions of the amplified sequence which are of importance and position (↑) of the G80A SNP.



The 5′-end of the sequence, from 152 onwards, has 61% GC content and at 64°C 60-70% exists in its helical form. However the 3'-end region between bases 50-152 remains 100% helical at the recommended 64°C due to the higher GC content (70%). The G80A polymorphism lies at position 151 (indicated by the arrow in figure 2.18), just within the first melting domain and explains why at 61 and 64°C it was not

detected. At 67°C the region above 152 bases is completely denatured and therefore unsuitable for detection of mutations or polymorphisms in this area.

Refining the temperature differences to 1°C the melt map patterns shown in figure 2.18 indicated that no single temperature was suitable for screening the whole of this fragment. A temperature of 64°C was unable to detect the G80A SNP but is still required to screen the region above 152 bases. As can be seen in figure 2.19, 65°C can be used to demonstrate the presence of the G80A heterozygotes.

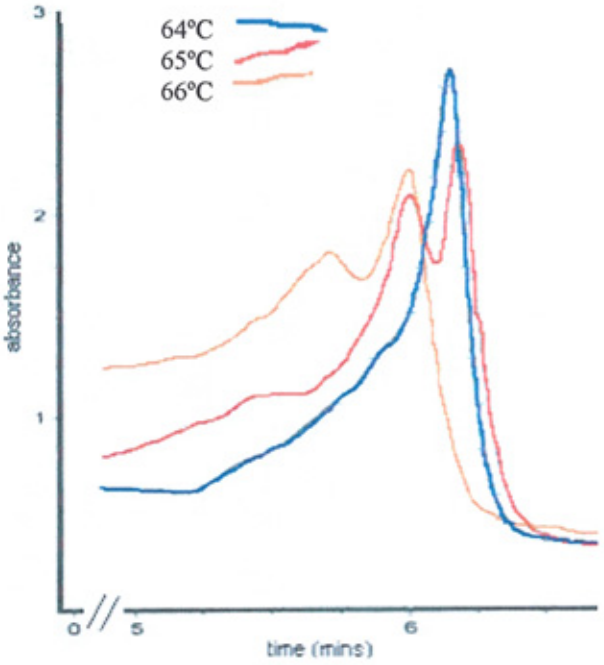


Figure 2.19 Superimposed heterduplex patterns generated by G80A heterozygous polymorphism at 64, 65 and 66°C.

Table 2.8 W AVE run conditions used for analysis of RFC exon 2 fragments at both 64 and 66°C.

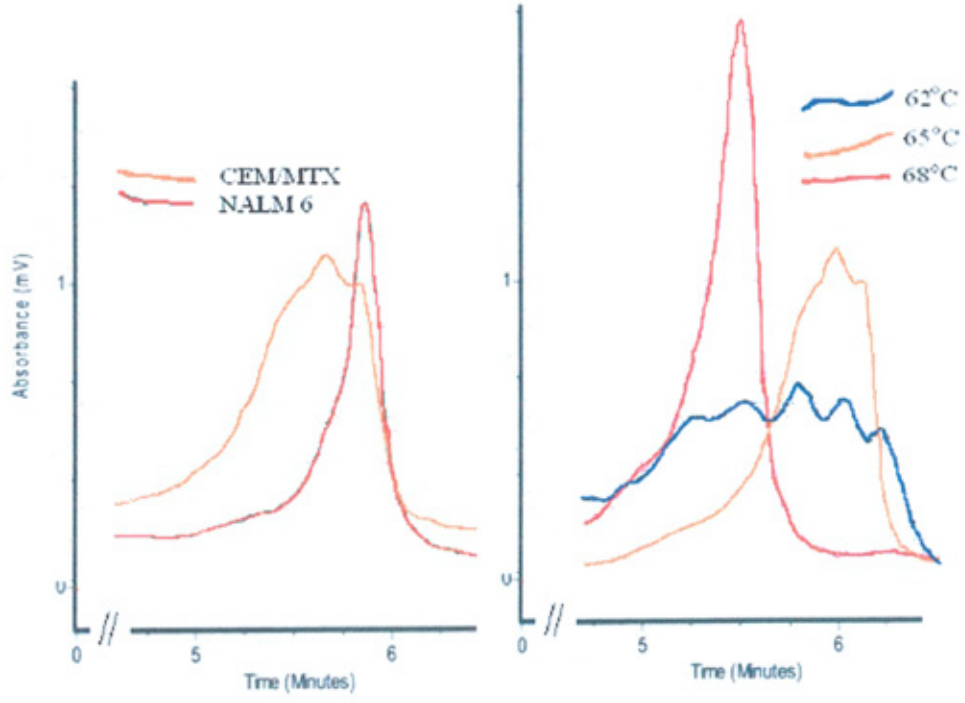
	Time (mins)		% A (0.1M TEAA)		% B (1.2M TEAA + 25% ACN)	
	64°C	66°C	64°C	66°C	64°C	66°C
Loading	0	0	54	50	46	50
Start gradient	0.5	0.5	49	47	51	53
Stop gradient	5.0	4.0	50	40	60	60
Start clean	5.1	4.9	0	0	100	100
Stop clean	5.6	5.1	0	0	100	100
Start equilibriate	5.7	6.5	54	50	46	50
Stop equilibriate	6.6	8.5	54	50	46	50

In this case, the optimum temperature calculated by the WAVEMAKER™ algorithm was the average for the whole fragment, but because of the extreme differences in GC content of the two domains, it proved unsuitable and therefore further work used 64, 65 and 66°C for WAVE® analysis, with 65°C proving the most useful for genotyping the G/A SNP.

2.5.7.3 **Optimisation of WAVE® conditions for analysis of exon 3 PCR products**

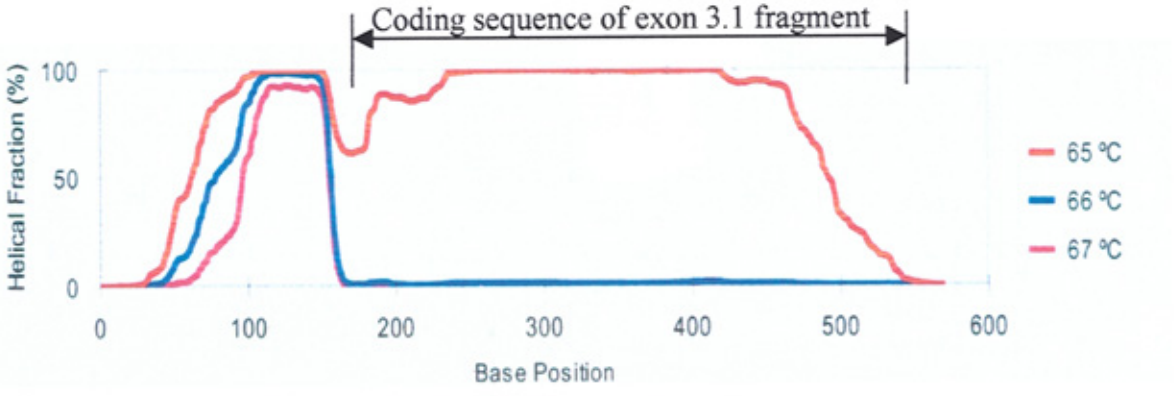
The CEM/MTX cell-line provided by Dr Jansen also contains a C352T mutation on one of the alleles which lies within the product from the RFC exon 3.1 pair of primers covering the 5'-end of this exon (Jansen *et al.*, 1998). The homoduplex pattern from NALM6 and the CEM/MTX heteroduplex pattern give clear discrimination at the WAVEMAKER™ predicted 65°C (figure 2.20). Following the experience with exon 2 it was decided to use the empirical approach and check at 62 and 68°C (figure 2.20).

Figure 2.20 (Left) Superimposed RFC exon 3.1 homoduplex (NALM6) and heterduplex (CEM/MTX) chromatogram profiles at 65°C. (Right) Superimposed profiles for CEM/MTX at 62, 65 and 68°C to demonstrate discrimination of heteroduplex pattern at different temperatures.



Before going ahead with analysis of normal and relapse samples, the melt map was scrutinised (figure 2.21). This melt map shows that the predicted temperature of 65°C is probably the best, although the 100-160 and 240-400 domains may not be screened effectively because they exist in a stable duplex form. Whilst the 100-160 domain can be selectively denatured at 67°C to give a 75-85% helical fraction, it was felt that no further temperature specific runs would be made because this domain does not contain the coding sequence of interest. The 240-400 domain is almost totally in the helical form and suggests that this relatively long sequence would be stable. However, the domain dramatically changes from 100% helical to total denaturation in just 1°C. This indicates the domain is not as stable as it first appears and that any mismatch in this apparently sensitive region should be detected. Overall, as the coding sequence for exon 3 starts at position 173 then using the parameters shown in table 2.9, screening at the predicted 65°C is considered suitable for analysis of the relapse and normal samples.

Figure 2.21. Melt map generated by WAVEMAKER™ software for the RFC exon 3.1 fragment showing the percentage of helical DNA at a given position for three temperatures (65, 66 and 67°C). Region of the amplified sequence which is of importance is shown by arrow and limit markers.



The 3'-end of exon 3 has more than one melting domain (see figure 2.22). Whilst the predicted temperature of 64.4°C covers most of the coding sequence, it was decided to run at 66.4°C as well to ensure the region between 100-280 bp was also covered. The WAVE® patterns at these temperatures for NALM6 provided clear homoduplex patterns (figure 2.23) and therefore the chromatography run parameters shown in table 2.9 were used for screening the normal and relapse samples.

Table 2.9 WAVE® conditions used for analysis of RFC exon 3.1 fragments.

	Time (mins)	% A (0.1M TEAA)	% B (1.2M TEAA + 25% CAN)
Loading	0	43	57
Start gradient	0.5	38	62
Stop gradient	5.0	29	71
Start clean	5.1	0	100
Stop clean	5.6	0	100
Start equilibriate	5.7	43	57
Stop equilibriate	6.6	43	57

Figure 2.22. Melt map generated by WAVEMAKER™ software for the RFC exon t3.2 fragment showing the percentage of helical DNA at a given position for three temperatures (64.4, 65.4 and 66.4°C). Region of the amplified sequence of importance is shown by arrow and limit markers. Position of the C696T polymorphism (rs12659) is shown

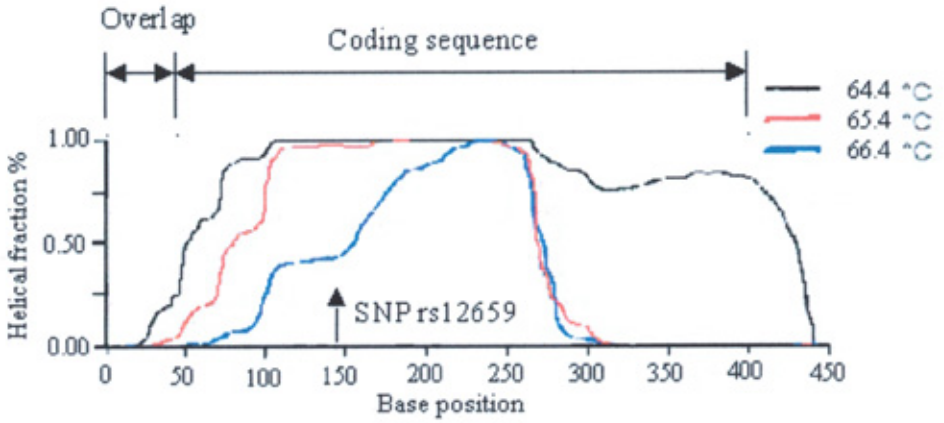
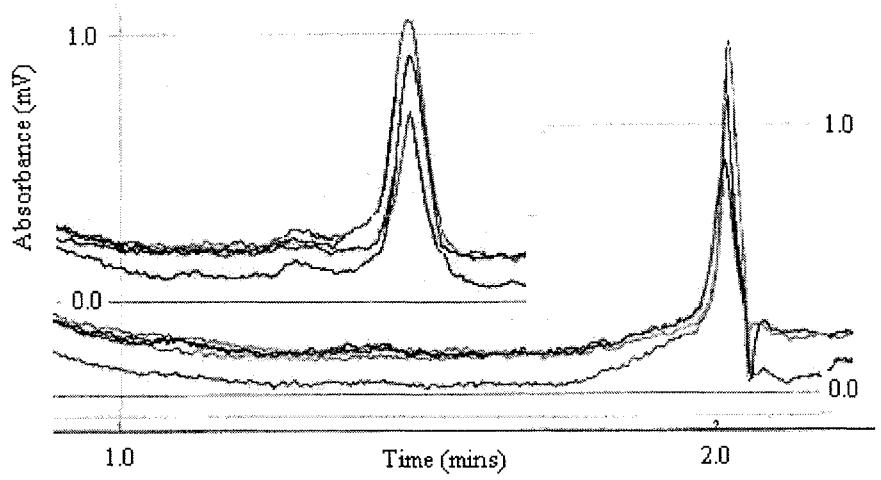


Table 2.10 W AVE conditions used for analysis of RFC exon t3.2 fragments.

	Time (mins)	% A (0.1M TEAA)	% B (1.2M TEAA + 25% ACN)
Loading	0	43.3	56.7
Start gradient	0.1	40.3	59.7
Stop gradient	2.1	30.3	69.7
Start clean	2.2	43.3	56.7
Stop clean	2.3	43.3	56.7
Start equilibriate	2.4	43.3	56.7
Stop equilibriate	2.5	43.3	56.7

Figure 2.23 Homoduplex patterns for NALM6 RFC exon t3.2 at 64.4°C and superimposed 66.4°C (upper left quadrant).



The relatively low PCR efficiency resulted in poor sensitivity of the chromatograms. Therefore the improved sensitivity provided by fluorescence detection of the WAVE®HS system was used and this generated the traces shown in Figure 2.24 (left). At first the shouldering and double peak at around 2.1 minutes was thought to indicate a true difference between homo and heteroduplex patterns, but there was no indication of these heteroduplexes from the UV traces of the same sample.

When the gradient table (table 2.10) was scrutinised, the anomaly at 2.1 minutes occurs at the end of the gradient and appears to be an artefact produced as the pumps and valves switch between delivery of the two buffers. Observation of pump

pressures during a run by the technical team confirmed that there is a brief surge of mobile phase at this point and this is likely to be the cause of the anomaly. Because the fluorimeter lies downstream of the UV detector there is no surge seen in the UV chromatogram as it had passed through the detector before 2.1 minutes, but is passing through the fluorimeter at this point. To confirm that this surge was the true cause of the heteroduplex split peak, a series of samples prepared for sequencing were analysed at 64.4°C using the WAVEMAKER™ predicted conditions and then run again using adjusted conditions to make the peak elute earlier (table 2.11). As can be seen in figure 2.24, the peaks were now all homoduplex and this adjusted programme was used for analysis of samples.

Figure 2.24. (Left) shows chromatographs produced from the fluorescent detector with apparent homoduplex traces (red) and heteroduplex patterns (blue). (Right) shows paired chromatograms of samples run at WAVE® predicted conditions and again with adjusted parameters so that the peaks are eluted earlier (2.1 and 1.6 minutes respectively)

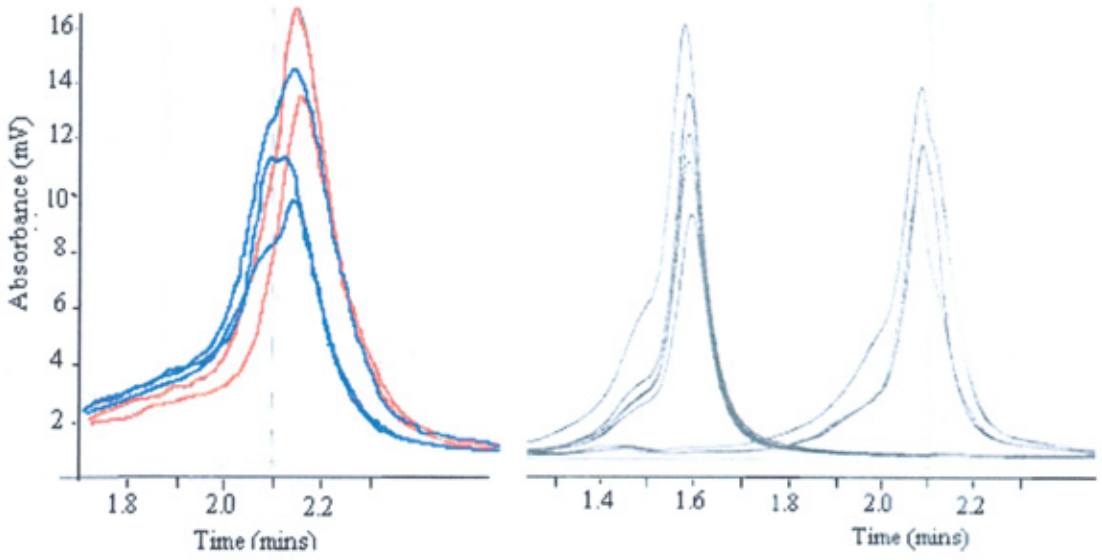


Table 2.11 WAVEMAKER™ predicted and adjusted gradient parameters to produce earlier elution of peaks for 64.4°C. A = 0.1M TEAA, B = 1.2M TEAA + 25% ACN

	Time (mins)	64.4°C % A	64.4°C % B	Adjusted 64.4°C % A	Adjusted 64.4°C % B
Loading	0	43.3	56.7	40.8	59.2
Start gradient	0.1	40.3	59.7	37.8	62.2
Stop gradient	2.1	30.3	69.7	27.8	72.2
Start clean	2.2	43.3	56.7	40.8	59.2
Stop clean	2.3	43.3	56.7	40.8	59.2
Start equilibriate	2.4	43.3	56.7	40.8	59.2
Stop equilibriate	2.5	43.3	56.7	40.8	59.2

2.5.7.4 Optimisation of WAVE® conditions for analysis of exon 4 PCR products

The melt map for the PCR product containing exon 4 has three domains (figure 2.25) but the WAVE® predicted temperature of 66°C appears to be suitable for screening. At this temperature the central domain is already partially denatured and the slightly more stable domains either side would be sensitive to any heteroduplexes, as they quickly go from 100% helical to <50% within 1°C. Chromatograms using run parameters shown in table 2.12 produced classic homoduplex patterns (figure 2.26).

Figure 2.25. Melt map generated by WAVEMAKER™ software for the RFC exon 4 fragment showing the percentage of helical DNA at a given position for three temperatures (66, 67 and 68°C). Region of the amplified sequence of importance is shown by limit markers.

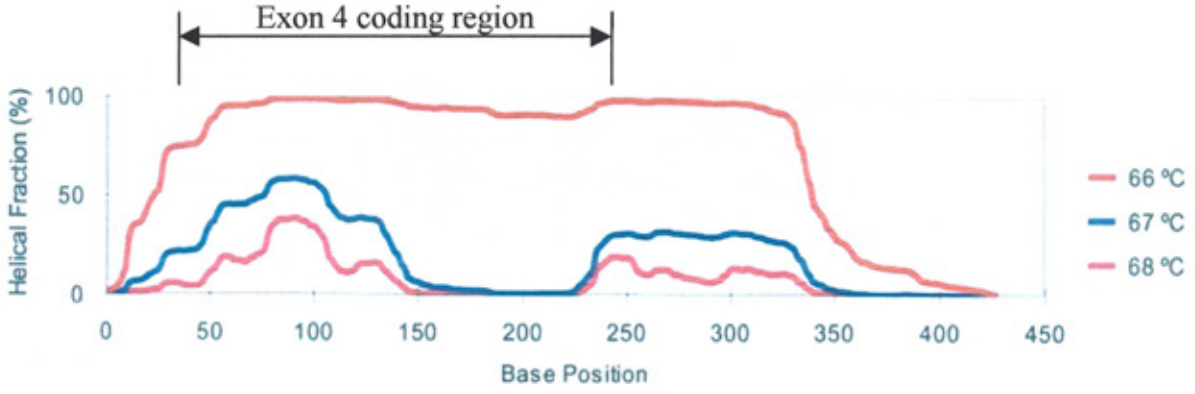


Figure 2.26. Classic homoduplex chromatograms obtained using WAVEMAKER™ predicted conditions at 66°C.

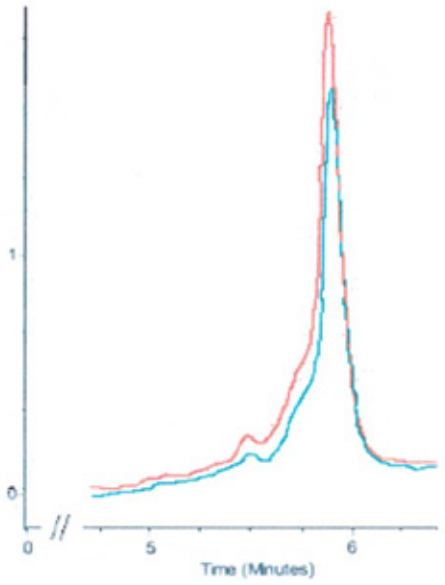


Table 2.12 W AVE conditions used for analysis of RFC exon 4 fragments.

	Time (mins)	% A (0.1M TEAA)	% B (1.2M TEAA + 25% ACN)
Loading	0	48	52
Start gradient	0.5	43	57
Stop gradient	5.0	34	66
Start clean	5.1	0	100
Stop clean	5.6	0	100
Start equilibriate	5.7	48	52
Stop equilibriate	6.6	48	52

2.5.7.5 **Optimisation of WAVE® conditions for analysis of exon 5 PCR products**

In the course of events, this exon was the first to be analysed. Without a suitable positive control to check that the conditions were acceptable and based on the initial homoduplex pattern seen on the WAVE® analysis (figure 2.27) then analysis of unknowns had gone ahead using parameters in table 2.13. However, as a result of problems with other exons the melt map for exon 5 was retrospectively scrutinised (figure 2.28). The GC rich domain between 75-100 bases probably requires WAVE®

analysis to be performed at 66°C as well as the predicted temperature of 62°C. However many of the samples were already depleted and/or unsuitable for reanalysis.

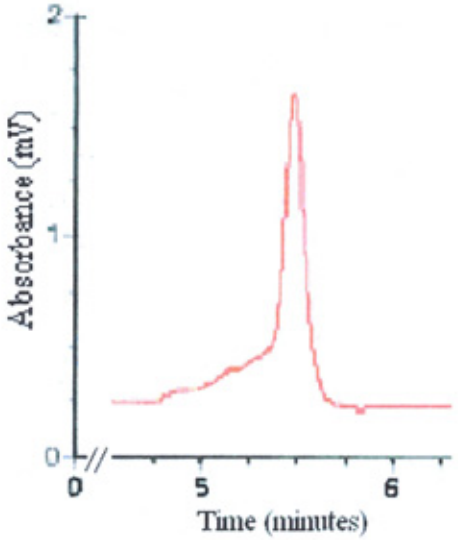
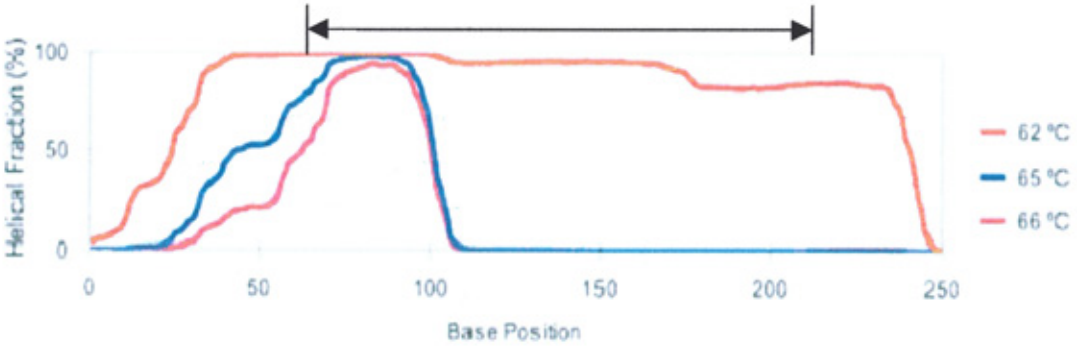


Figure 2.27. Homoduplex pattern produced by NALM6 for RFC exon 5.

Table 2.13 W AVE conditions used for analysis of RFC exon 5 fragments.

	Time (mins)	% A (0.1M TEAA)	% B (1.2M TEAA + 25% ACN)
Loading	0	52	48
Start gradient	0.5	47	53
Stop gradient	5.0	38	62
Start clean	5.1	0	100
Stop clean	5.6	0	100
Start equilibriate	5.7	52	48
Stop equilibriate	6.6	52	48

Figure 2.28. Melt map generated by WAVEMAKER™ software for the RFC exon 5. showing the predicted 62°C temperature, plus 65 and 66°C to determine when the GC rich region between 75-100 bp started to denature. The exon sequence itself is shown by limit markers.



2.5.7.6 **Optimisation of WAVE® conditions for analysis of exon 6 PCR products**

The WAVEMAKER™ predicted temperature of 62.5°C may be the temperature at which 75% of the sequence is helical, but scrutiny of the melt map (figure 2.29) showed the presence of significantly different melting domains that make analysis at a single temperature unsuitable. The coding sequence lies between 129 and 612 bp and includes two adjacent domains with extremes in T_m. At 60.8°C the region between 130 and 180 bp has a lower melting temperature than the predicted temperature for the whole sequence, whilst the region from 200-400 remains 100% helical at the predicted temperature. Increasing the temperature sequentially by 1°C had little effect on this region, even up to 66°C. At this point, an alternative graphical format available on NAVIGATOR™ software for analysing the sequence and determining the optimum T_m was used (figure 2.30). From this, it appears a temperature as high as 69°C is required to achieve 75% in the helical form for the short sequence around the 250 bp position that is 90% GC rich.

Figure 2.29. Melt map generated by WAVEMAKER™ software for the RFC exon 6. showing the predicted 62.5°C temperature, plus 60.8 and 63.5°C. The important coding sequence is shown by arrow and limit markers.

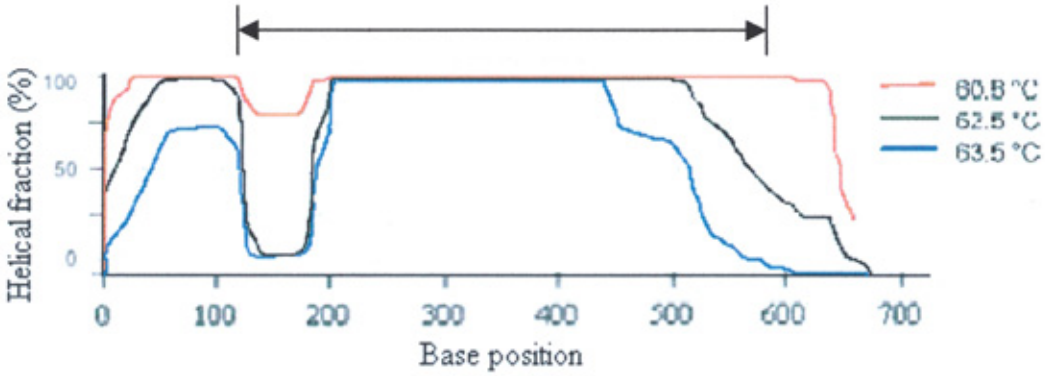
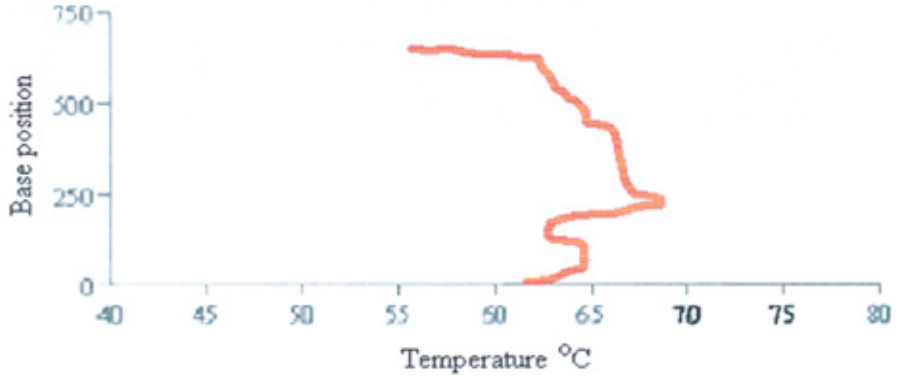


Figure 2.30. Alternative melt map format generated by WAVEMAKER™ software for the RFC exon 6 showing base position against the temperature required to render 75% in helical form at that position.

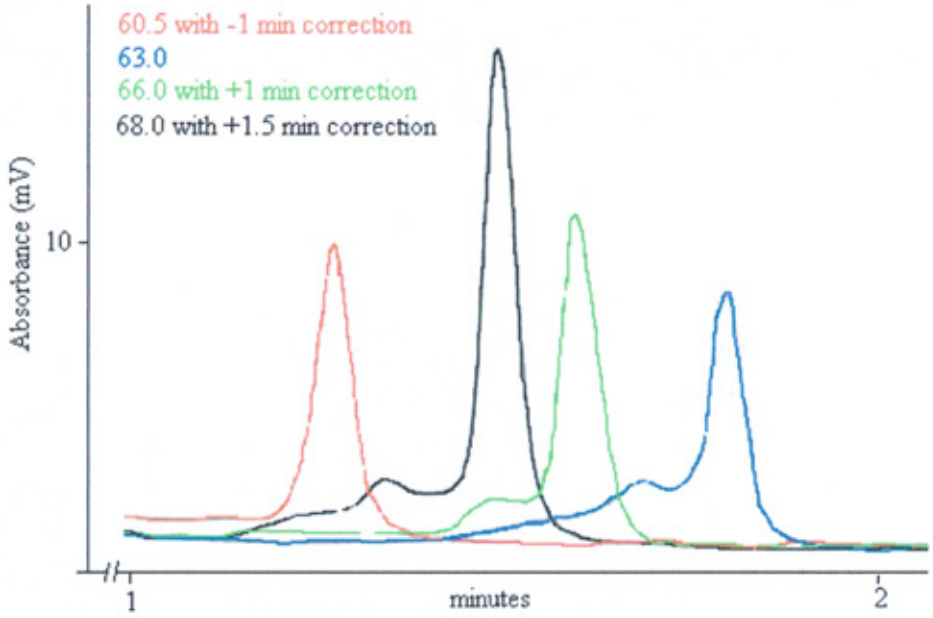


It was clear that a single temperature was not going to be suitable, so a series of three pools (NALM6, K562 and CEM/MTX) were run at a range of temperatures to determine what the chromatograms looked like. A temperature of 60°C failed to produce a peak, probably because it was being retained on the column too long, whilst 66 and 68°C eluted too early and were difficult to resolve from the peak made up of non-retained material. Using time-adjusted parameters for the gradient (table 2.14) the peaks were eluted within the 1.0 – 2.0 minute time window and four of these 60.5, 63.0, 66.0 and 68.0°C were selected for analysis of the unknown samples (figure 2.31).

Table 2.14. Adjusted gradient conditions to compensate for run temperature of RFC exon p6 products. Where A = 0.1M TEAA and B = 1.2M TEAA + 25% acetonitrile

		60.5°C		61.5°C		62.5,63.5,64.5 and 65.5°C		66.5 and 67.5°C		68.5°C	
	Time (mins)	%A	%B	%A	%B	%A	%B	%A	%B	%A	%B
Loading	0.0	36.0	64.0	38.5	61.5	41.0	59.0	46.0	54.0	48.5	51.3
Start gradient	0.1	33.0	67.0	35.5	64.5	38.0	62.0	43.0	57.0	45.5	54.5
Stop gradient	2.1	23.0	77.0	25.5	74.5	28.0	72.0	33.0	67.0	35.5	64.5
Start clean	2.2	36.0	64.0	38.5	61.5	41.0	59.0	46.0	54.0	48.5	51.5
Stop clean	2.3	36.0	64.0	38.5	61.5	41.0	59.0	46.0	54.0	48.5	51.5
Start equilib	2.4	36.0	64.0	38.5	61.5	41.0	59.0	46.0	54.0	48.5	51.5
Stop equilib	2.5	36.0	64.0	38.5	61.5	41.0	59.0	46.0	54.0	48.5	51.5

Figure 2.31. Homoduplex pattern produced by NALM6 for RFC exon 6 at the four temperatures chosen for further analysis of unknown samples.



2.5.8 ANALYSIS OF NORMAL AND RELAPSE SAMPLES

2.5.8.1 **Panel of normal gDNA**

The Northern Institute for Cancer Research had acquired a panel of gDNA as part of the North Cumbria Community Genetics Project (NCCGP) in accordance with ethical approval WCRLEC no347 and was available for this project. The DNA was obtained and prepared as described in section 2.5.8.3 from either maternal peripheral or foetal cord blood and remained anonymous. This finite resource was used over the period 2001-2006 by several researchers in the group and as a result DNA from many of the wells was depleted before completion of this project, preventing full profiles of all exons to be completed.

2.5.8.2 **Panel of relapse gDNA**

DNA was prepared as detailed in section 2.5.8.3 from samples collected routinely during diagnostic and therapeutic procedures for patients referred to the Royal Victoria Infirmary, Newcastle-upon-Tyne with childhood ALL between 1988 – 2002.

Samples were collected and used in accordance with local ethical approval guidelines (reference 2002/111). Diagnosis and disease status at relapse were confirmed using FAB classification, immunophenotyping, cytogenetics, white cell count and patients subsequently allocated a treatment protocol. If they had been available, paired presentation and relapse samples would have been preferred, but such pairs were rarely available to the group. Accepting the limitations of DNA availability in childhood ALL, the strategy adopted was to initially screen relapse samples, as these will contain any acquired mutations. If a mutation or novel SNP was detected in the relapse sample, then the bank of presentation DNA was checked for the matching presentation sample. If available, this was analysed to determine if the mutation existed before treatment started. Originally, 40 relapse samples were selected from those available based on volume of sample and DNA concentration. Of these 29 were male with an average age at presentation of 6.03 years (range 1.5 – 11.4 years); 2 T-cell, 25 B-cell, 1 mixed lineage and 1 null classification. Eleven females with an average age at presentation of 8.06 years (range 0.8 – 14.1 years) were categorised as 3 T-cell, 5 B-cell and 1 mixed lineage. Of these 40 original samples only 26 had full analysis performed across all exons due to depleted gDNA before the study was completed.

2.5.8.3 **Preparation of genomic DNA.**

Genomic DNA (gDNA) was prepared using the QIAamp mini kit (Product code 51104; QIAGEN, Crawley, UK) following the manufacturer's instructions; 5 x 10^6 saline washed cells in a 2 mL Eppendorf tube were resuspended in 200 µL of phosphate buffered saline (PBS). 20 µL proteinase K, 4 µL RNase A and 200 µL buffer AL were added, mixed for 15 seconds and then incubated at 56°C for 10 minutes to release the DNA and digest interfering RNA. The tube was then briefly

centrifuged using the pulse spin function of the Jouan A14 microcentrifuge to bring all the sample to the bottom of the tube before 200 µL of 100% ethanol (AR grade) was added. Samples were then vortex mixed for 15 seconds and the mixture transferred to a QIAGen spin column. The column and capture tube were centrifuged for 1 minute at 8,000 rpm in the microcentrifuge to capture the DNA on the column filter. The column was then transferred to a fresh capture tube and 500 µL of buffer AW1 added to the spin column. The column and capture tube were again centrifuged at 8,000 rpm for 1 minute and the column then transferred to another fresh capture tube. 500 µL of buffer AW2 was added and the column and capture tube centrifuged for 3 minutes at 14,000 rpm in the microcentrifuge. The flow through was discarded from the capture tube and the column and empty capture tube respun for a further 3 minutes at 14,000 rpm. This ensures any remaining alcohol or wash buffer has been removed from the column. Finally the column is transferred to a 2 mL sterile Eppendorf tube and 200 µL of buffer AE carefully added to the filter of the column and left to stand for at least 1 minute. This releases the DNA from the filter which is then eluted into the collection tube by centrifugation at 8,000 rpm for 1 minute.

The purity and concentration of the gDNA preparation was then determined by measuring the absorbance at 260 and 280 nm against water in a UV suitable 96-well microtitre plate on the UV microtitre plate reader (Costar, High Wycombe, UK). The ratio of absorbances at 260/280 nm indicates purity of the DNA, where a value close to 1.8 is considered high purity. Values greater or lower than this indicate protein impurities. Whilst the cell populations processed in this way principally contained the leukaemic lineage, there was the possibility that a proportion of normal cells would also be included and could influence the DHPLC patterns.

2.6 RESULTS

2.6.1 Scoring of WAVE® chromatograms

Some of the early experiments generated low PCR product concentration, and whilst the peak could be seen on the chromatogram, the low absorbance values required expansion of the Y-axis and as a result there was amplification of the background noise making interpretation more difficult. For this reason, many months were spent modifying and repeatedly performing PCR on these samples until sufficient DNA was generated to give a signal >1 mV and provide confidence in the interpretation of the chromatogram. This problem was overcome to some degree with the use of the WAVE®HS system introduced in 2003, which has the option to enhance the signal by the use of its fluorescent detection system.

Performance of the WAVE® was checked in each run using a mutation standard that produces a classic four peak heteroduplex pattern (Product code 562000; Transgenomic Inc, San Jose, CA, USA) and allows inter-run performance to be checked. A positive control, (ie., one with a known SNP or mutation), was used whenever possible, to confirm that the settings for each run were optimal. In addition, a negative and wild-type control was used for the exon being analysed. The negative control was prepared at the PCR stage with a well or tube containing all reagents except DNA. This was used to check for non-specific amplification at the PCR stage and the sample was then used on the WAVE® to confirm a stable base line. Each of the test chromatograms was compared against the homoduplex wild-type pattern generated by the NALM6 sample which was also set up in each PCR batch and subsequently analysed in a WAVE® run. Comparison was greatly enhanced with the introduction of NAVIGATOR™ software that allowed normalisation of each of the chromatograms from one or more runs. Normalisation is the mathematical

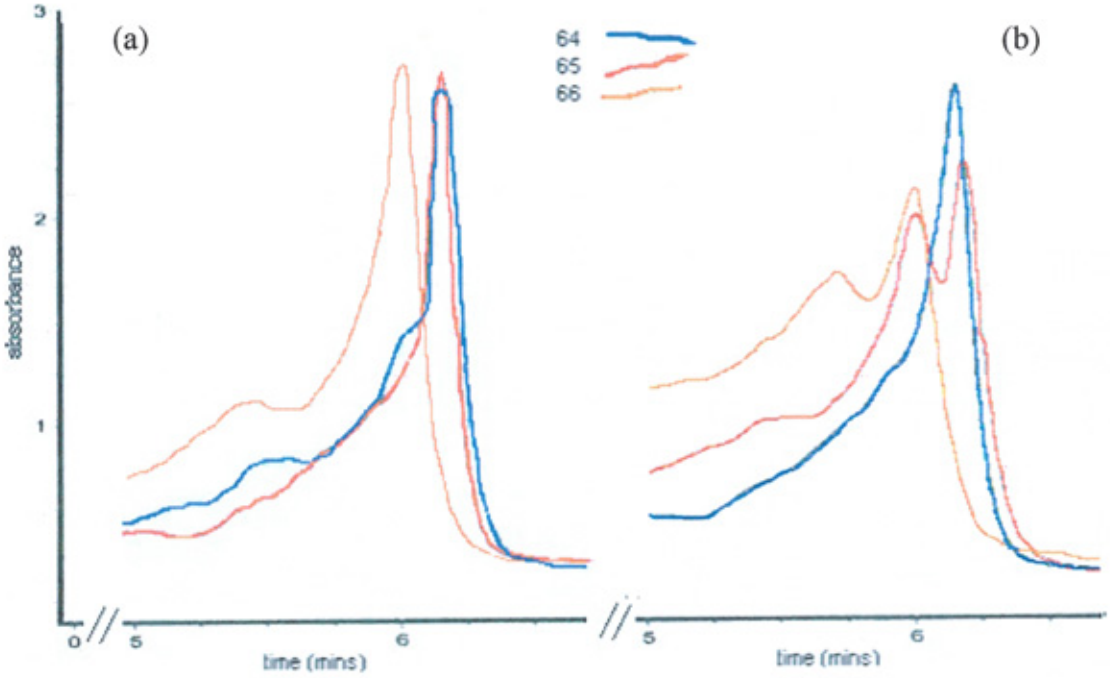
manipulation of the chromatogram so that they each have the same sensitivity on the Y-axis and any slight analytical variation in retention time experienced in any chromatographic procedure is eliminated.

2.6.2 RFC exon 2

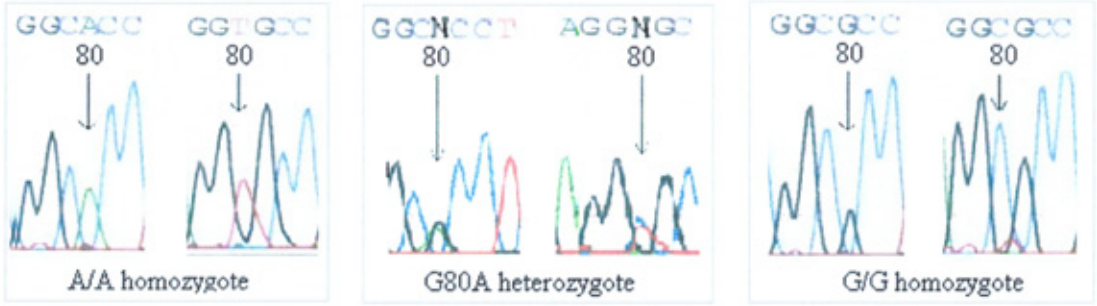
Three clearly different chromatogram patterns were identified. The most common are those shown in panels (a) and (b) of figure 2.32. Sequencing of representative samples with these patterns demonstrated that the heteroduplex pattern seen in panel (b) was heterozygous for the G80A polymorphism. NALM6 shown in panel (a) gave a clear homoduplex pattern which when sequenced was confirmed as the G/G homozygous genotype (see Appendix 2.2). NALM6 was selected for spiking of homozygous samples as described in section 2.2. Samples which continue to have a homoduplex pattern following spiking and repeat DHPLC analysis were classified as homozygous G/G genotype, whilst those which now produced a heteroduplex pattern must have been homozygous A/A and were classified as such. Initially, scoring of homo- and heteroduplex patterns was by visual appearance, but to reduce observer-biased interpretation, a ratio of the area under the curve for the shouldered or early peak (5.9 – 6.0 minutes) against the main peak (6.1 minutes) at 65°C was calculated. A ratio of <1.0 was used to classify the chromatogram as a homoduplex, whilst >1.0 indicated a heteroduplex.

CEM/MTX DNA failed to produce an atypical pattern expected of the positive control. Repeat PCR and DHPLC analysis continued to show a pattern typical of the G80A genotype and yet subsequent sequence analysis confirmed the presence of the expected G/A mutation at position 133 (numbered from the start codon and shown in panel (d) of figure 2.32).

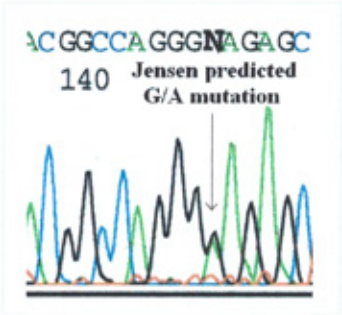
Figure 2.32. Chromatogram patterns produced from RFC exon 2. Panel (a) shows homoduplex patterns for NALM6 at 64, 65 and 66°C. Panel (b) shows the most frequently seen heteroduplex pattern at the same three temperatures and represents the heterozygous G80A polymorphism. Panel (c) shows the forward and reverse sequence data at position 80 to confirm each of the three genotypes. Panel (d) shows the G133A mutation in the CEM/MTX cell-line.



(c)



(d)



Considering the report by Kaufman *et al* in 2004 that RFC mutations are unlikely to be the mechanism for MTX resistance, it was no surprise that of the 44 relapse samples analysed only one atypical chromatogram pattern that could not be attributed to a known SNP was seen (sample 6250 shown in figure 2.33). PCR and chromatography were repeated on three occasions for this sample and each time confirmed the original result. This atypical pattern was therefore considered a significant finding and followed up by sequencing that confirmed a C to T transition at position -37, (ie., upstream of the start codon). The difference in sequence can clearly be seen in figure 2.34 and is taken to validate the analytical protocol used here to screen for unknown mutations. Whilst this mutation does not lie within the coding region of exon 2 and therefore has no direct effect on the protein sequence it may have an effect on the transcription of the gene or mRNA stability and will be discussed further in section 2.7.4.

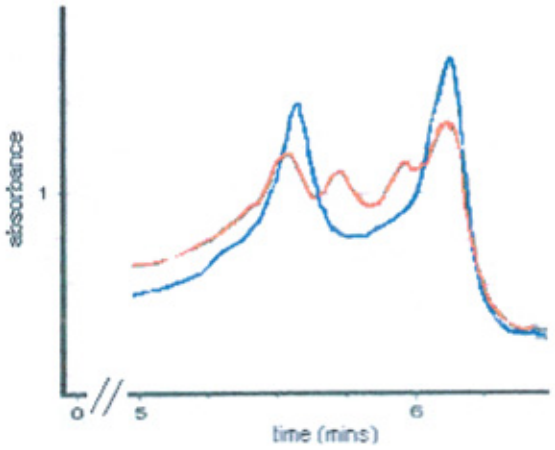
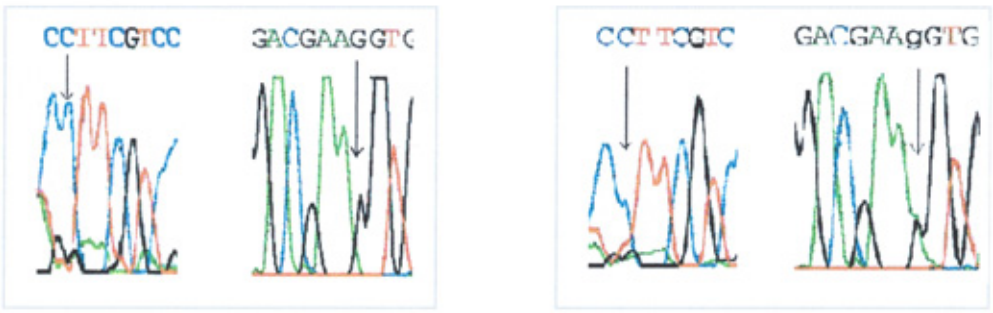


Figure 2.33 Atypical chromatogram patterns at 64 (blue) and 65°C (red) for relapse sample 6250.

Figure 2.34 Forward and reverse sequencing data for normal (left panel) and relapse sample 6250(right) showing C-37C in normal and C-37T in relapse sample.



Three further samples were identified that had been taken from patient 6250 at different points in the patient's history and were amplified and analysed by DHPLC and sequencing. A sample collected at original presentation did not show evidence of the mutation, nor did a sample when this patient unfortunately first relapsed. However, a second relapse did show the change identified in the latter sample 6250 (3^{rd} relapse) and indicated that this mutation likely occurred between the first and second relapse. The chronological development of the mutation is an interesting finding that will be discussed further in section 2.7.4.

Full results for normal and relapse samples are tabulated in appendix 2.3. Of the 81 normal samples successfully genotyped at 65°C 38.3% (n = 31) were scored as G/G; 49.4% (n = 40) as G/A and 12.3% (n = 10) as A/A. The 44 relapse samples were scored as 31.8% (n = 14); 56.8% (n = 25) and 11.4% (n = 5) respectively. χ^2 analysis for determining the significance of differences in distribution of results across groups was applied using the statistical function in Excel (Microsoft Corp. Seattle, USA). Analysis (see appendix 2.4) demonstrated that with a P value >0.5, there is no significant difference between the distribution of genotype between normal and relapse samples.

2.6.3 RFC exon 3

The CEM/MTX cell line generated a heteroduplex pattern (figure 2.20) as expected using the 5′-end primers, but when the sample was sequenced it was not the C258T reported by Jansen (1998) (shown on sequence in appendix 3.1). Neither was it due to the NCBI SNP Cluster ID: rs1051269 reported to produce a silent mutation at Pro82Pro (G246C). The only difference identified when aligned with NALM6 was a G to A transition in the intronic region upstream of the exon 3 coding sequence

(figure 2.35). This appears to be the likely cause of the heteroduplex pattern. With the exception of one normal sample (N77), the normal or relapse samples only generated homoduplex patterns. Sample N77 produced the double peak pattern shown in figure 2.36, but there was insufficient DNA to repeat the PCR, chromatography and sequencing to confirm this finding and thus identify the genotype causing this pattern.

Figure 2.35 Forward and reverse sequencing data for NALM6 (left panel) and CEM/MTX (right) showing the only difference identified when both sequences from the RFC exon 3.1 primers were aligned.

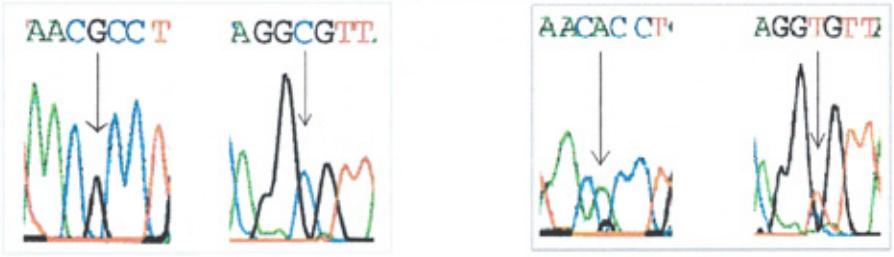
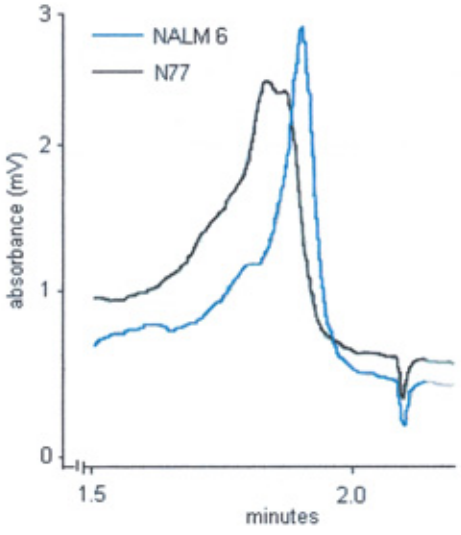


Figure 2.36 Heteroduplex pattern produced by normal sample N77, compared with NALM6 homoduplex pattern



WAVE® analysis of the 3′-end of the coding sequence for exon 3 (primers t3.2) produced a relatively consistent heteroduplex pattern (figure 2.37; left). Whilst the difference was not great at any of the three temperatures used, the appearance of a small pre-peak was clearly seen and considered suitable for scoring the sample. When the normalise tool, available on NAVIGATOR™ was used, the difference is clearly seen (figure 2.37; right). Sequencing of selected samples with homoduplex and heteroduplex patterns (both before and after spiking) demonstrated that the differences were attributable to the C696T SNP previously reported on the NCBI database as rs12659 (figures 2.38) and not the rs9282853 silent SNP.

Figure 2.37 (Left) Typical heteroduplex patterns compared to the homoduplex pattern of NALM6, note sample 1775 was run on two separate occasions. (Right) Normalised chromatograms for two homoduplex and two heteroduplex patterns.

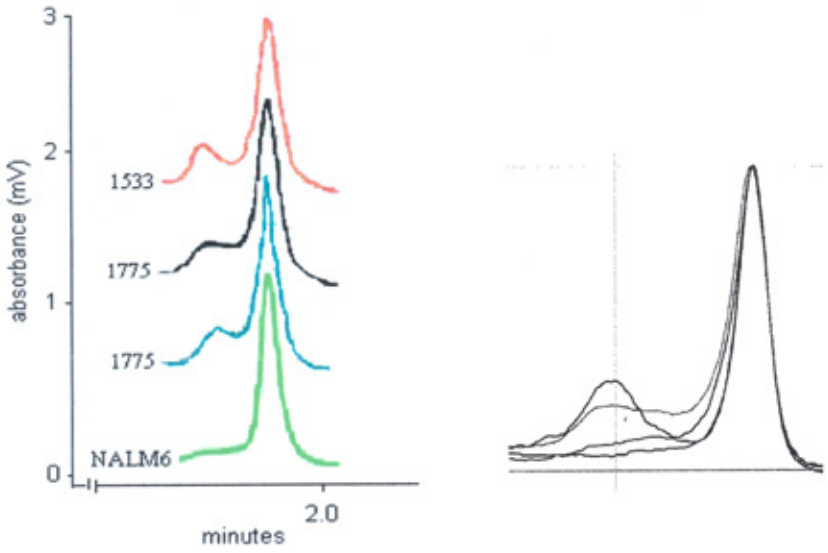
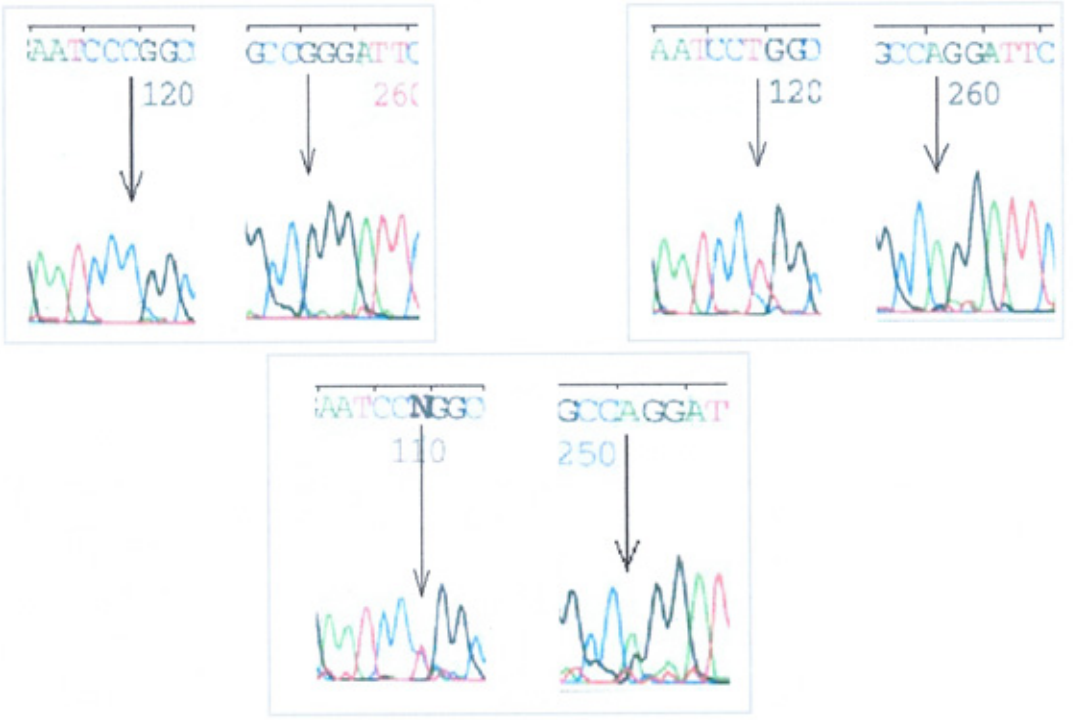


Figure 2.38 Examples of each of the genotypes for the C696T SNP; (top left C696C; top right T696T; bottom C696T (samples N63, 5467 and 1775 respectively)



Of the few papers that report a genotype frequency for the C696T polymorphism, Merola *et al* (2002) propose a relatively equal distribution of the genotypes (30.0% C/C; 37.5% C/T and 32.5% T/T), whilst Kaufman *et al* (2004) show a clear bias to the C/C genotype (85% C/C; 14% C/T and only 0.4% T/T). From the small sample size in this study the C allele has even stronger penetrance (95.3% C/C; 3.5% C/T and 1.2% T/T) with no significant difference ($0.10>P>0.05$) between the normal and relapse samples using χ^2 (see appendix 3.8).

2.6.4 RFC exon 4

Initial screening generated three samples with heteroduplex patterns (figure 2.39). Sequencing of these samples gave poor results from the forward primer and alignment of the reverse sequences failed to show any difference in normal (N) sample N25 or relapse sample 3605. Repeat analysis of N25 was unable to replicate the pattern and

therefore the original heteroduplex pattern was considered an artefact. There was insufficient gDNA to repeat analysis of relapse sample 3605. Sample N36 produced a likely G972A heterozygous sequence (figure 2.40 – note only the reverse sequence is shown) which was identified as the only difference when aligned against the other three sequenced samples and that expected from the wild type chromosome 21 sequence.

Figure 2.39 Heteroduplex patterns seen in normal samples N25 and N36 and relapse sample 3605 compared to homoduplex pattern produced by NALM6

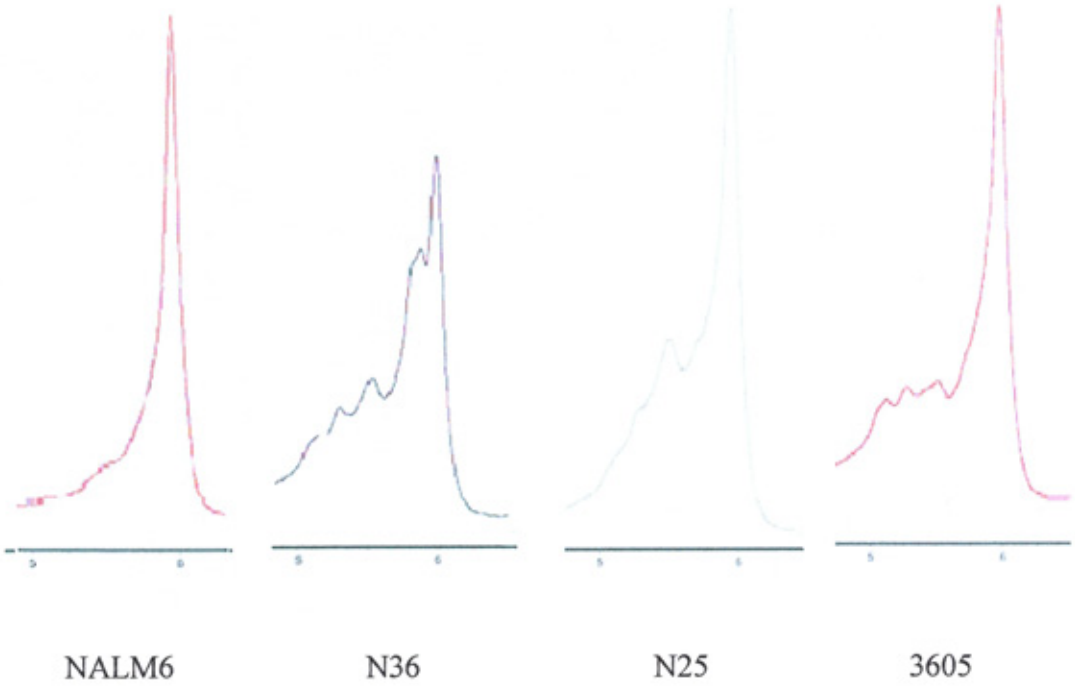
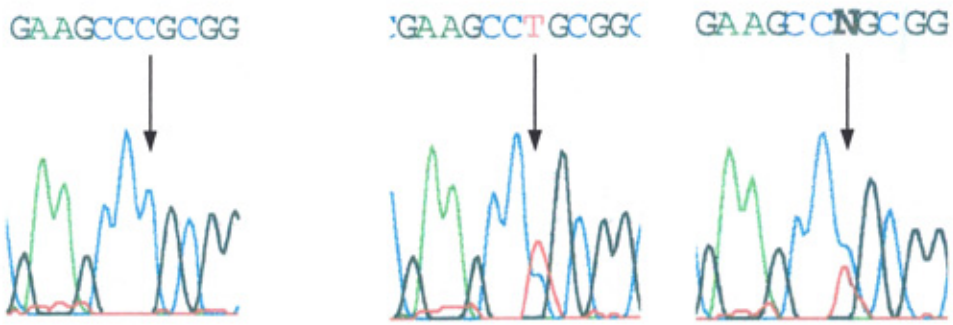


Figure 2.40 Sequence data from RFC exon 4 reverse primer showing wild type homozygous pattern (G972G) for codon 324 (left) and heterozygous patterns from sample N36 (G/A) on two separate occasions (centre and right). NOTE:- these are the reverse sequences.



Kaufman *et al* (2004) report this silent mutation (Ala324Ala) as a SNP with frequencies of 98% for G/G; 2% for G/A; which compare well with the frequencies obtained here (97% G/G; 2% G/A and 1% A/A). Although the latter needs to be confirmed by sequencing, it can be concluded that the homozygous A972A would appear to be at a very low frequency.

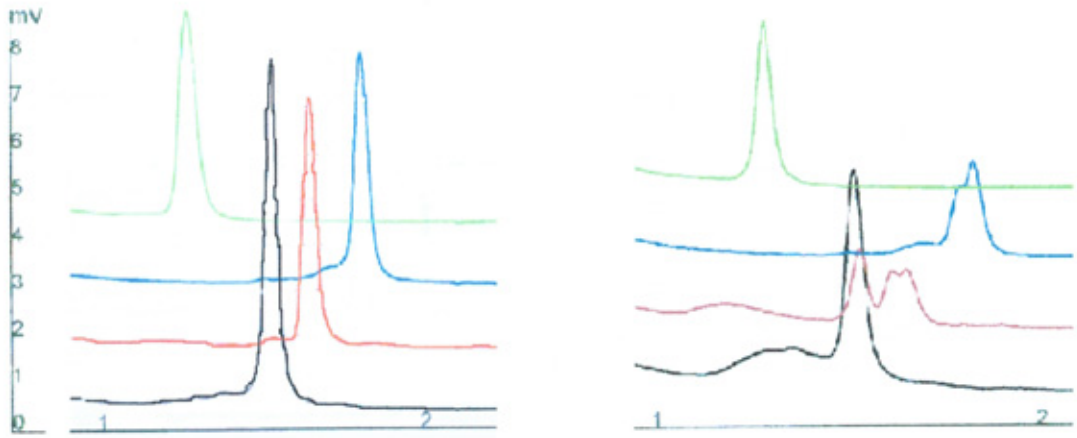
2.6.5 RFC exon 5

All chromatograms obtained were identical and thus taken as wild type sequences identical to NALM6 (appendix 5.2).

2.6.6 RFC exon 6

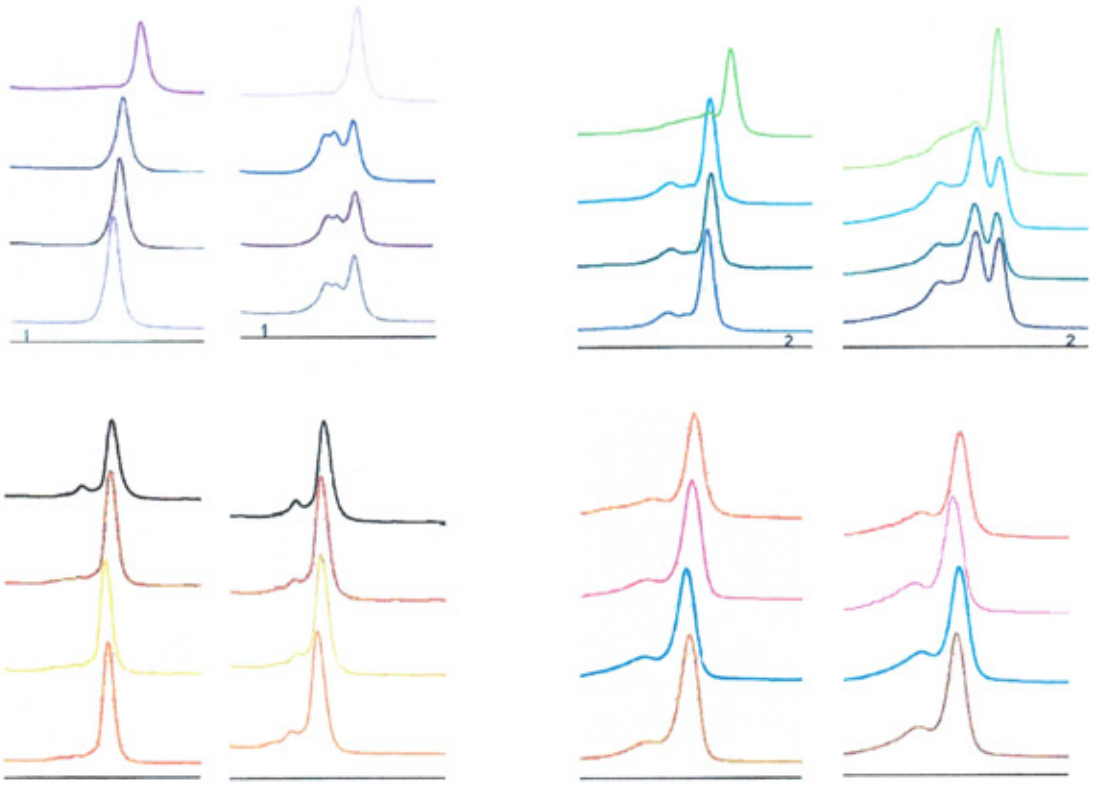
Sample N63 gave heteroduplex patterns at 63.0 and 66.0°C (see figure 2.41). Unfortunately there was insufficient sample left to sequence and identify the genomic difference.

Figure 2.41. Chromatograms for sample N63 showing heteroduplex patterns at 63°C (blue) and 66°C (red) and homoduplex patterns at 60.5°C (green) and 68°C (black). For comparison the homoduplex patterns obtained for NALM6 are shown on the left.



All other normals and relapse samples were scored as homoduplex at all four temperatures. However on spiking, samples 7096, N36 and N50 gave clear heteroduplex patterns (figure 2.42). These samples were purified and sequenced but no obvious difference could be seen to account for this (sequence data not shown). All three samples were analysed in a single small batch and therefore this may be an artefact. To confirm the original findings analysis should have been repeated, but unfortunately there was insufficient material to repeat the amplification and subsequent WAVE® analysis on any of these three samples and verify the heteroduplex formation when spiked.

Figure 2.42. RFC exon 6 chromatograms obtained for four samples before and after spiking with NALM6 (wild-type) at 60.5 (top left), 63.0 (top right), 66.0 (bottom left) and 68.0°C (bottom right). In each set of four traces, NALM6 is top, then in descending order, sample 7096, N36 and N50.



2.7 DISCUSSION

2.7.1 Overcoming problems in PCR amplification.

The GC content of the RFC exons (table 2.15) generally reflected the difficulty experienced in their amplification. Exon 5 was selected for familiarisation of the techniques and initial experiments because it was the smallest. Successful primer design and PCR amplification at the first attempt for this exon gave a false impression of the practical work to follow on the rest of the gene. Looking back, this may have been not because of its size but because this exon has a GC content of only 57%. The problems encountered in amplification of GC rich exons were not unexpected, as others have reported similar problems with other genes (Agarwal and Perl, 1993; Weissensteiner and Lanchbury, 1996), but the time spent in trying to resolve the problems was greater than anticipated.

Table 2.15. GC and AT content of RFC exon PCR products (exon 3 is split into two fragments for amplification each with similar GC content as the coding sequence).

	Coding sequence		PCR product	
	%AT	%GC	%AT	%GC
Exon 2	34	66	34	66
Exon 3	33	67	33	67
			34	66
Exon 4	32	68	30	70
Exon 5	44	56	43	57
Exon 6	38	62	34	66

The problem arises because the three hydrogen bonds involved in the GC pairing of aligned dsDNA make the pairing naturally stronger than the two bonds between A and T. In GC-rich regions the strength of a single pairing is further enhanced by GC pairs either side of it (see section 2.2), so that a higher melting temperature is required

for denaturation. GC-rich dsDNA is also more likely to form strong secondary structures that further resist denaturation at the predicted temperature (Mytelka and Chamberlain, 1996) and structures such as hair-pin loops hinder access and annealing of primers. Such secondary structures are likely to occur toward the 3′-end of a gene (Ljungman and Hanawalt, 1992). Whilst they are causing problems for *in-vitro* analysis it would be reasonable to assume they have evolved in nature to slow *in-vivo* DNA polymerase progression at points where fidelity is more critical. Retarded polymerase activity i*n-vitro* can result in PCR fragments shorter than expected because of insufficient time for complete extension (Daniel and Mytelka, 1996) and may account for some of the problems encountered in PCR.

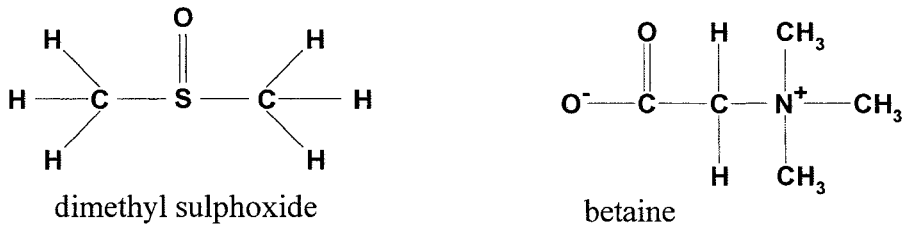
Forcing the amplification of a GC-rich region by simply increasing reagent concentration and lowering the annealing temperature can result in non-specific binding and errors in the product (Daniel and Mytelka, 1996), so alternative solutions which retain fidelity were explored. The most successful approach has been the addition of destabilising agents.

Modification of salt concentrations normally present in the buffer mixture facilitated a degree of optimization without adding molecules that could potentially alter the chromatographic separation, or worse, irreversibly damage the column. Magnesium binds and neutralises the phosphate backbone of DNA and in effect facilitates helix formation. Reducing the concentration of $MgCl_2$ did have some beneficial effect for exons 2 and t3.2 amplification as the binding affinity between the two strands of DNA was lowered and a lower T_m required. This effect however is non-specific and as concentration is reduced it no longer supports DNA polymerase activity and negates any benefit from destablising the dsDNA.

The more specific targeting by alkyl ammonium salts such as tetramethyl-ammonium chloride (TMAC) and tetraethylammonium chloride (TEAC) (Shapiro *et al.,* 1969) was considered. These salts are relatively small strong electrolytes that readily dissolve and gain access to the base pairs via the major groove where they preferentially bind to the AT pairing and interfere with the hydrophobic stacking energies and render AT and GC bonds the same (isostabilisation). However, unless there is a clean-up procedure to subsequently remove these salts from the PCR product, then this approach was not considered because the salts would affect the ion-pairing buffers essential for successful DHPLC. Other destabilization agents which have proved effective elsewhere, were not tried because they have not been cleared by Transgenomic for safe use on the WAVE® and include nucleotide analogs (Jenkins and Turner, 1995; Mutter and Boynton 1995), formamide (Sarkar *et al.,* 1990), glycerol, trehalose, 2-pyrrolidone (Chakrabarti and Schutt, 2001a). Use of such agents would invalidate the warranty on the column. DMSO and Betaine have been validated for use on the WAVE®.

DMSO (figure 2.43) was found to enhance amplification to some degree in almost all PCRs. At low concentrations DMSO does not affect *Taq* DNA polymerase activity (Gelford, 1989; Cheng *et al.,* 1994; Varadaraj and Skinner, 1994) and can, in fact, improve efficiency and fidelity. These properties were utilised to improve the effectiveness of Optimase™ used in amplification of exon 6. It disrupts base pairing by redistributing electrons and decreases inter- and intra-strand reannealing before primers can locate the complementary sequence on the ssDNA (Frackman *et al.,* 1998; Chakrabarti and Schutt, 2001b).

Figure 2.43 Diagram showing simple structures of DMSO (left) and betaine (right).



Betaine ($N_1N_1N_1$-trimethylglycine or (carboxymethyl)-trimethylammonium) (figure 2.43) is a modified glycine normally used by the cell as a methyl donor. It has been used for several decades as an osmoprotectant and has several other *in-vitro* benefits to the molecular biologist (Weissensteiner and Lanchbury, 1996; Hengen, 1997). It is less viscous than glycerol, cheap and non-toxic (Melchior and von Hippell, 1973). Its success in improving PCR amplification of exons 3 (using primer pair t3.2) and 4 is a result of it binding to AT pairs via the major groove and increasing the hydration of GC pairs as it binds and attracts water via the minor groove (Mytelka and Chamberlain, 1996; Hengen, 1997). Increasing the hydration of the GC pairs affects the hydrophobicity established by the van der Waals interaction between stacked bases. These interactions are weak by themselves, but in dsDNA provide stability. Therefore allowing access of water to the bases reduces this stability and the energy required to denature dsDNA is lowered. The overall effect is an equilibrium of the T_m required to denature GC and AT pairing down toward the T_m of an AT rich region (isostabilisation). Hengen (1997) demonstrated that betaine allowed the denaturation temperature in PCR to be dropped to 92-93°C and the primer annealing temperatures reduced by 1-2°C. Whilst amplification is improved, Hengen (1997) also cautioned against the possibility of increased misincorporation when lowering the temperature. Weissensteiner and Lanchbury (1996) suggested that betaine provided greater tolerance of PCR conditions (eg., $MgCl_2$ concentration no longer critical) as an

alternative theory on how betaine improved amplification, Rees *et al* (1993) that it was a result of improved polymerase stability during the high temperature denaturation step in each cycle and Henke *et al* (1997) that it reduces secondary structure formation in the dsDNA and thus facilitates polymerase procession.

Whatever the mechanism, it was clear that for these two exons, and as will be seen in chapter 3, several of the DHFR exons, betaine significantly enhanced PCR amplification and facilitated DHPLC analysis of the products. The use of betaine was preferred in this work because its zwitterionic structure meant that it should have minimal effect on the charges required for the ion-pairing properties of DHPLC separation. This was tested on NALM6 and K562 for two exons, and thankfully no difference was noted in the chromatogram patterns or retention times.

2.7.2 Effectiveness of screening using DHPLC

The classic four peak heteroduplex pattern that would have made scoring of samples easy was seen for the mutation identified in exon 2 (ie., sample 6250, figure 2.31), but was elusive in screening for other heteroduplex products. This was not unexpected as others have reported difficulties in interpreting chromatograms (Xiao & Oefner, 2001; Pfeiffer *et al.,* 2002) and attribute the variation in heteroduplex patterns seen to the base composition of the remaining helical sequence, near neighbour sequence, degree of hydrogen bonding between mismatched pairs and even the DNA polymerase used (Tombline *et al.,* 1996; Liu *et al.,* 2002; Speigelman *et al.,* 2000). Kosaki *et al* (2005) have suggested that impure primers which have a percentage with a single base missing can generate products which will differ in size by a single base and this may not resolve completely into a separate peak, but can cause confusing peak broadening or shouldering. Although not confirmed as the cause, such band broadening was seen

during the project and required repeat amplification of several sample runs. The fact that gDNA prepared from the relapse samples may have contained a proportion of normal cells, should also be considered to have an effect on the chromatogram patterns. Xiao & Oefner (2001) also report that mutations and SNPs in GC-rich regions cause peak broadening. With several of the RFC exons being GC-rich this was considered during interpretation of the chromatograms.

The problems in interpretation of atypical patterns have been partially resolved by the use of the normalisation function available in the WAVEMAKER™ software. The ability to compare suspect peaks with reference peaks (eg., NALM6) by software that makes allowance for differences in the intensity of absorbance or fluorescence signal and analytical variation in retention time, proved to be very useful. The fluorescent detection capability of the newer WAVE®HS allowed interpretation of samples that previously would have been discarded with insufficient signal (Baharami *et al.,* 2002). Both of these improvements remove some of the human variation (Cotton and Bray, 2001), and were utilised here and enhanced confidence in the results obtained. These improvements have been used elsewhere to increase screening throughput by analysis of sample pools (Kuklin *et al.,* 1999; Giordano *et al.,* 2001; Frueh and Noyer-Weidner, 2003) or multiplex PCR of single samples for several sequences (Xiao and Oefner, 2001; Dehainault *et al.,* 2004). The improved sensitivity has also enhanced the detection of MRD down to <0.5% (Stadt *et al.,* 2001; Frueh and Noyer-Weidner, 2003). DHPLC is so precise that it can be used for genotyping of known SNPs without the need for sequencing (Underhill *et al.,* 1997; Gross *et al.,* 2000; Ezzeldin *et al.,* 2002; Frueh and Noyer-Weidner, 2003). This greatly enhances the power of DHPLC as a screening tool and was used to score the G80A, C696T and G972A SNPs.

2.7.3 Single nucleotide polyporphisms in the RFC gene

The importance of single nucleotide polymorphisms is only just becoming clearer as published research demonstrates the link between genotype and disease, or genotype and response to treatment. Because of the direct effect on the resulting protein the majority of interest is focused on SNPs in the exons. However, there is an increasing interest in the intronic regions and both 3'- and 5'-UTRs because of the influence they can have on transcription factor control of a gene several kilobases along the chromosome.

The G80A SNP in exon 2 of the reduced folate carrier has been studied extensively in relation to gender (Pei *et al.,* 2004), race (Rady *et al.,* 2001), circulating folate levels (Chango *et al.,* 2000) and pathological conditions such as hyperhomocysteinaemia (Whetstine *et al.,* 2001b; Hirakoa *et al.,* 2004), thrombotic risk (Yates and Lucock, 2005) and congenital defects (DeMarco *et al.,* 2001; Shaw *et al.,* 2002, 2003; Morin *et al.,* 2003; Pei *et al.,* 2004). The genotype frequency demonstrated in this current work, although a relatively small study, is comparable with that published elsewhere (see appendix 2.5).

The only other reported SNP causing a sequence change in RFC to be aware of is rs7278825 (shown in appendix 6.1). The C1401T polymorphism causes alanine at residue 469 to change to valine and this could alter the configuration of the C-terminal loop, but work by Ng and Henikoff, (2001) suggest that this change will be tolerated. Silent SNPs in the RFC-1 gene (rs1051269, rs9282853, rs12659), have been documented (Merola *et al.,* 2002; Kaufman *et al.,* 2004; Packer *et al.,* 2006). Whilst they may be of use as DHPLC quality controls, there is no data to suggest that one or

more polymorphisms in the non-coding sequence of the RFC gene alone would confer any resistance leading to relapse of the patient group compared to the normal group.

2.7.4 Mutations in the RFC gene (relapse sample 6250)

A lot of knowledge about the reduced folate carrier has been gained from mutations in a range of malignant cell lines (see chapter 1). The relative ease with which mutations in the RFC-1 gene were introduced to cell lines *in vitro* and the frequent presence of sequence alterations to this protein in osteosarcoma, suggested that *in vivo* changes in ALL would be likely (Gorlick *et al.,* 1997). It was therefore disappointing (from a research point of view) that only one somatic mutation was found in this study. Whilst disappointing, this reflected the findings of Kaufman *et al* in 2003, that the most likely cause of MTX resistance involving RFC is due to decreased expression, rather than changes in the protein structure. Assaraf (2007) reviewed potential mechanisms for silencing of the RFC-1 gene at the level of (a) transcription factor defects; (b) CpG island promoter methylation status; or (c) changes in gene copy number. These mechanisms may coexist in order to present with resistance to antifolates (Kaufman *et al.,* 2006).

Although there is a lack of somatic mutations in the RFC-1 gene, the WAVE® proved very successful in detecting the heteroduplex pattern generated by the relapse sample 6250 and thus fulfilled one of the aims of this thesis, to screen for somatic mutations. Repeat PCR, DHPLC and sequencing of sample 6250 confirmed the presence of a novel mutation in exon 2, not within the coding sequence as hoped, but upstream of the start codon (shown in appendix 2.1), potentially in a region that influences transcription factor binding and therefore expression.

Sample 6250 was one of four samples from a girl who had unfortunately relapsed on three occasions. The individual presented on the 5th July 1990, aged 3.5 years with a common B-cell lineage. Cytogenetics were recorded as 46 XX/55, +X, +6, +17, +17 incomplete. Analysis of gDNA from the presentation sample and a remission sample collected on 13th August 1992 were normal for exon 2 (WAVE® analysis confirmed by sequencing). The patient relapsed for the first time in November 1992, but no sample was available. gDNA prepared from a bone marrow taken at second relapse in January 1994 was also analysed and the C-37T mutation was confirmed as present. Sample 6250 was gDNA collected at a subsequent third relapse in November 1994. This suggests that the mutation is not constitutive, but a true mutation which arose between remission in 1992 and is likely to have contributed to relapse (second relapse if not the first relapse) and unfortunately the child's death in February 1995.

To understand what role it may play, a search was performed using one of the functions of the Omega software package to look for known motifs in this region of the gene and determine if any have been lost or introduced as a result of the mutation. No existing motifs in the wild-type sequence were recognised and therefore the loss of a transcription binding site is thought to be unlikely. Altering the cytosine to thymine produces a sequence recognised as the BK virus promoter on the lead strand. The BK virus promoter is one of the oncogenic DNA viruses and infects up to 80% of humans by adulthood (Kraus *et al.,* 2001). The virus has been associated with an increased risk of certain cancers (Reploeg *et al.,* 2001; Weinreb *et al.,* 2006), but this requires the involvement of the whole virus and not just the introduction of the promoter. Steroid/thyroid/retinoid hormone receptors will bind to the virus promoter, blocking its transcription (ie suppression) (Schreier and Gruber, 1986). Therefore, if this site exists in the promoter region of the RFC-1 gene it is possible that

steroid/thyroid/retinoid hormone receptors bind to it and block the usual upstream transcription control of exon 1. Evidence would need to be produced to support or discredit this theory.

BK virus promoter acagcgtcaC**T**TTCgtcccctccgg on the lead strand
tgtcgcagtG**A**AAGcaggggaggcc

The CpG methylation site is too far upstream of the RFC-1 coding region to include the C-37T mutation and thus have a recognised epigenetic effect on the gene (Worm *et al.,* 2001). This supports the findings of Rothem *et al* (2004) and Kastrup *et al* (2007) that changes in methylation status of the RFC promoter region do not cause the decreased expression in ALL seen at presentation or relapse.

2.7.5 **Evidence of a mechanism to confer MTX resistance in sample 6250**

The most likely cause of resistance involving RFC is an alteration to expression of the protein (Gorlick *et al.,* 1997; Trippett and Bertino, 1999). To investigate if this was related to mRNA stability being modified by the mutation, a set of primers were designed to recognise RFC mRNA using the Primo Pro 3.4 software (*http://www.changbioscience.com/primo/primo.html*).

RFC mRNA forward GAA ACT CCT GTC CTG GGG AG
RFC mRNA reverse GCG CCA TGA AGC CGT AGA AG

The product is sufficiently different in size from that generated from the gDNA sequence (which will be 3714 base pairs in length because of the intronic sequence). The primers now need to be tested on a sample containing human mRNA to determine if they work and then use them to determine if the levels of mRNA for RFC-1 are abnormal. This work is still to be completed.

2.8 RFC Summary

Denaturing HPLC was shown to be a useful tool for the rapid screening of unknown mutations and polymorphisms in gDNA. The main problems were in the design of primers and optimisation of PCR for GC-rich regions of the genome. Running samples at the WAVE® predicted temperature was not always appropriate due to multiple melting domains and several temperatures had to be used. Whilst no novel somatic mutations in the coding sequence were identified, the previously unreported mutation at -37 in exon 2 may have some effect on the transcription of the gene or stability of the resulting mRNA. Table 2.16 shows the summary findings of WAVE® analysis for each exon, including the genotyping of known SNPs.

Table 2.16 DHPLC scoring of RFC exons. Samples with complete profiles of exon 2-6 are shaded. Where possible genotypes of known SNPs are shown.

Sample number	EXON 2 G80A SNP	EXON 3.1	EXON 3.2 C696T SNP	EXON 4 G972A SNP	EXON 5	EXON 6
L68					W/T	
76	G/A	WT	C/C	G/G	W/T	W/T
171	G/A	WT	C/C	G/G	W/T	W/T
279	G/G	WT	C/C	G/G	W/T	W/T
L308		WT	C/C	G/G		
L345		WT	C/C	G/G		
L382		WT	C/C	G/G		
405	G/A	WT	C/C	G/G	W/T	W/T
L609	G/G	heteroduplex	C/C	? G/A		
980	G/A	WT	C/C	G/G	W/T	
981				G/G	W/T	
1399	G/G	WT	C/C	G/G	W/T	W/T
1533	G/A	WT	C/T	G/G	W/T	W/T
1566	G/A	WT	C/C	G/G	W/T	
1720	G/G	WT	C/C	G/G	W/T	W/T
1775	G/A		C/T	G/G	W/T	W/T
1785	G/G	WT	C/C	G/G	W/T	W/T
1920						W/T
2045	G/G	WT			W/T	
2416	A/A	WT	C/C	G/G	W/T	W/T
2426	G/A	WT	C/C	G/G	W/T	W/T
2578	G/A	WT	C/C	G/G	W/T	W/T
2650	G/A	WT	C/C	G/G	W/T	W/T
2865	G/A	WT	C/C	G/G	W/T	W/T

Table 2.16 continued

Sample number	EXON 2 G80A SNP	EXON 3.1	EXON 3.2 C696T SNP	EXON 4 G972A SNP	EXON 5	EXON 6
3016						W/T
3192	G/A	WT	C/C	G/G	W/T	W/T
3451	A/A	WT	C/C	G/G	W/T	W/T
3605	G/A	WT	C/C	G/G	W/T	W/T
3990	A/A		C/C	G/G	W/T	W/T
4225	A/A	WT	C/C	G/G	W/T	
4684	G/G	WT	C/C	G/G	W/T	W/T
4845	G/G	WT			W/T	
4900	G/A	WT	C/C	G/G	W/T	W/T
5467	G/A	WT	T/T	G/G	W/T	W/T
5836	A/A	WT	C/C	G/G	W/T	W/T
5862(27/9)	G/A	WT		G/G	W/T	
5887	G/G	WT	C/C		W/T	
6250	G80G C-37A	WT	C/C	G/G	W/T	W/T
6346	G/G	WT		G/G		
6534		WT		G/G		
6563	G/A	WT		G/G	W/T	W/T
6652	G/A	WT	C/C	G/G	W/T	W/T
6845						W/T
7096	G/A	WT	C/C	G/G	W/T	?
8646	G/A	WT	C/C	G/G	W/T	W/T
8668	G/A	WT	C/C	G/G	W/T	W/T
8678	G/A	WT	C/C	G/G	W/T	W/T
8679	G/A	WT		G/G	W/T	W/T
8743						W/T
8839	G/G	WT	C/C		W/T	W/T
8871	G/G		C/C	G/G	W/T	
9212			C/C			W/T
9216	G/A	WT		G/G	W/T	W/T
9506	G/G	WT	C/C	G/G	W/T	W/T
11420	G/A			G/G	W/T	

N1	G/A	WT		G/G	W/T	W/T
N2	G/A				W/T	
N3	G/A			G/G	W/T	
N4	G/G				W/T	
N5	G/G	WT	C/C	G/G	W/T	W/T
N6	G/A				W/T	
N7	G/G	WT		G/G		
N8	G/G	WT	C/C	G/G	W/T	W/T
N9	G/A				W/T	
N10					W/T	
N11	G/A	WT	C/C		W/T	
N12	G/G			?A/A	W/T	
N13	G/G	WT			W/T	
N14	A/A	WT			W/T	
N15	G/G		C/C	G/G	W/T	W/T
N16	A/A	WT	C/C		W/T	
N17	A/A	WT	C/C	G/G	W/T	

Table 2.16 continued

Sample number	EXON 2 G80A SNP	EXON 3.1	EXON 3.2 C696T SNP	EXON 4 G972A SNP	EXON 5	EXON 6
N18	G/A	WT	C/C	G/G	W/T	W/T
N19	A/A				W/T	
N20	G/A	WT	C/C	G/G		
N21	A/A	WT	C/C	G/G	W/T	W/T
N22	G/A	WT			W/T	
N23	G/A	WT	C/C		W/T	W/T
N24	G/G	WT	C/C		W/T	
N25	G/G			G/G	W/T	
N26	G/A	WT	C/C	G/G	W/T	W/T
N27	G/A	WT	C/C	G/G	W/T	W/T
N28	G/A	WT	C/C	G/G	W/T	W/T
N29	G/G	WT	C/C	G/G	W/T	W/T
N30	G/A	WT	C/C	G/G	W/T	W/T
N31	G/G	WT	C/T	G/G		
N32	G/A	WT	C/C	G/G	W/T	W/T
N33	G/A			G/G	W/T	
N34	G/G			G/G	W/T	
N35	A/A				W/T	
N36	G/G	WT	C/C	G/G	W/T	?
N37	G/A	WT		G/G	W/T	
N38	G/G	WT	C/C	G/A	W/T	W/T
N39	G/A	WT	C/C	G/G	W/T	W/T
N40	G/G	WT	C/C	G/G	W/T	W/T
N41	G/G	WT				
N42	G/A					
N43	G/A					
N44	G/G					
N45	G/A	WT				
N46	G/A					
N47	G/A					
N48	G/A					
N49	G/A	WT	C/C	G/G	W/T	W/T
N50	G/G	WT	C/C	G/G	W/T	?
N51	G/G	WT	C/C	G/G	W/T	
N52	G/A	WT	C/C	G/G	W/T	W/T
N53	G/A					
N54	G/A	WT				
N55	G/G					
N56	G/G	WT				
N57	G/G	WT	C/C	G/G	W/T	W/T
N58	G/A	WT				
N59	G/G	WT				
N60	G/A	WT	C/C	G/G	W/T	
N61	G/G					
N62	G/A					
N63	G/G	WT	C/C	G/G	W/T	
N64	G/A	WT	C/C	G/G	W/T	
N65	G/G	WT	C/C	G/G	W/T	W/T
N66	G/A					

Table 2.16 continued

Sample number	EXON 2 G80A SNP	EXON 3.1	EXON 3.2 C696T SNP	EXON 4 G972A SNP	EXON 5	EXON 6
N67	G/A					
N68	A/A	WT		G/G	W/T	W/T
N69	G/A	WT	C/C	G/G	W/T	W/T
N70	A/A	WT	C/C	G/G	W/T	
N71	G/G	WT	C/C	G/G	W/T	W/T
N72	A/A	WT	C/C	G/G	W/T	
N73	G/G		C/C		W/T	
N74	G/A		C/C	G/G	W/T	
N75	G/A				W/T	
N76	G/G	WT	C/C	G/G	W/T	
N77	G/A		C/C	G/G	W/T	
N78	G/G	WT	C/C	G/G	W/T	W/T
N79	G/A	WT	C/C	G/G	W/T	W/T
N80	G/A	WT	C/C	G/G	W/T	
N81	G/G	WT	C/C	G/G	W/T	W/T
N83	A/A	WT	C/C	G/G	W/T	W/T

3 SCREENING THE DIHYDROFOLATE REDUCTASE GENE

The second aim of this research was to screen the dihydrofolate reductase (DHFR) gene for somatic mutations or polymorphisms in the coding sequences that could provide a degree of MTX resistance and thus contribute to relapse. The methods applied to the mutational screening of RFC were applied to the DHFR gene in the same patient groups. To meet this aim the following objectives were established:-

a) Set-up and optimise PCR amplification of coding exons in the DHFR gene.
b) Optimise the denaturing high performance liquid chromatography screening method for each of the exons.
c) Amplify and screen a panel of gDNA available from children with ALL collected at relapse (sample most likely to have MTX related mutations) and compare this with gDNA from a panel of normal individuals

3.1 AMPLIFICATION OF THE DHFR GENE

In order to use DHPLC for screening gDNA, specific exons were amplified using PCR to produce sufficient material for detection and generate products of optimal size for separation. The following sections describe the design of primers for each of the DHFR exons, optimisation of PCR, optimisation of DHPLC and interpretation of the results generated by the WAVE® DHPLC system.

3.1.1 Primer design and optimisation of PCR conditions

Studies on DHFR in acute lymphoblastic leukaemia (ALL) have focused on mRNA expression or cDNA sequences (Volm *et al.,* 1994; Levy *et al.,* 2003) and as a result the primers from these papers may amplify processed pseudogenes in the DNA and produce artefact heteroduplex chromatograms. Therefore, in this study gDNA will be

amplified and analysed. None of the literature published up to late 2003 provided primer details for gDNA, so the first task at that time was to design primers and optimise PCR for the six coding exons.

Sequences from the EMBL database (ENSE00000307796) were used to identify the coding exons (2-7), and these were then aligned to the 5q11.2–13.2 region of chromosome 5, so that sufficient sequence either side of the exon could be selected and used for primer design. Any of the primers considered suitable for use were entered into the NCBI BLAST search facility (*http://www.ncbi.nlm.nih.gov/blast/*) to check that the primers did not align elsewhere on the genome and potentially cause non-specific product amplification. In addition, the NCBI SNP database (*http://www.ncbi.nlm.nih.gov/sites/entrez*) was searched for DHFR to check for potential SNPs that could affect primer recognition and annealing. Primers selected did not contain a known polymorphism.

All primers were ordered from MWG Oligo Synthesis-Bitoech AG (Ebersberg, Germany), stored and prepared for use as described in section 2.5.3 and appendix 1 A4. DNA polymerases and PCR reagents were prepared as described in section 2.5.4. Initial PCR for all primer pairs used the manufacturer's recommended reagent concentrations and thermal cycler conditions for 'touchdown' amplification as shown in table 2.3 and figure 2.7. These were subsequently modified for each exon as described in sections 3.1.2 – 3.1.7 using procedures established in the amplification and analysis of the RFC gene in chapter 2.

3.1.2 Design of exon 2 primers and optimisation of PCR

DHFR Exon 2m for	CGC CTG CAC AAA TGG GGA CGA G
DHFR Exon 2m rev	CCG CGC AGC AGA AAA GGG GAA TC

The pair shown above were generated by DNAStar and gave a 345 bp product with a predicted annealing temperature of 66°C differing by 1.1°C. However, a BLAST search for these primers not only indicated alignment with the target DHFR, but also sequences associated with the MSH3 gene (mutS homologue 3). This gene resides on chromosome 5 (Matheson *et al*, 2007) and when the two genes are aligned they show some overlap (figure 3.1). This explains why the primer is associated with another gene, but in fact, is highlighting the same region of the genome and will not generate an alternative product.

Figure 3.1. Alignment of DHFR and MSH3 genes on opposite strands of chromosome 5 showing 80 bp overlap (taken from Ensembl database ENST00000307796).

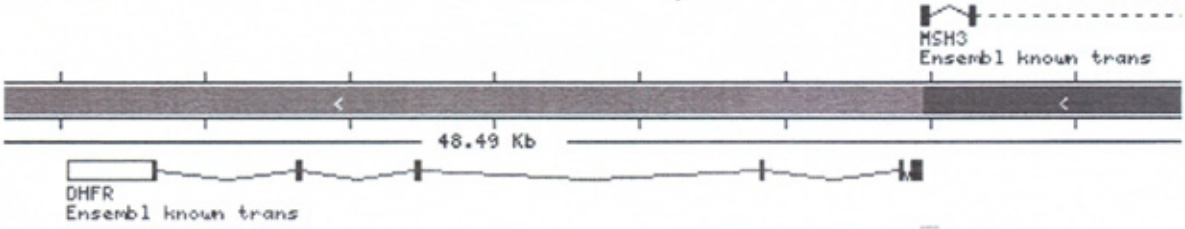
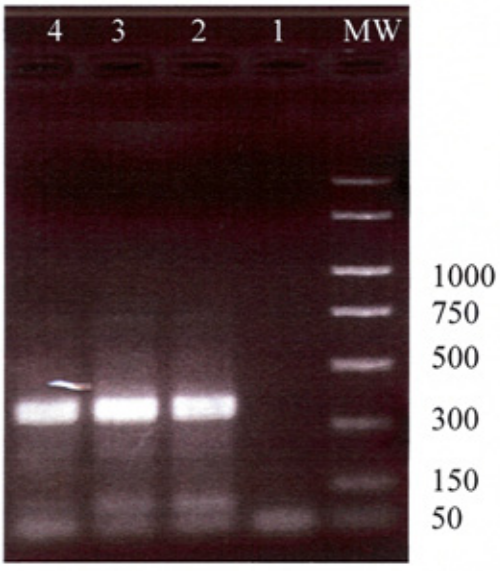


Figure 3.2. The gel shows products generated from the DHFR m2 primer pair using Optimase DNA polymerase; lane 1 = negative control, lanes 2 and 3 using touchdown PCR at 58°C, lane 4 = touchdown PCR at 65°C.

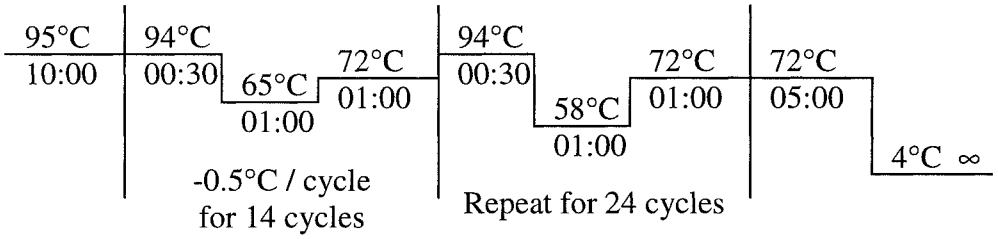


Following optimisation, a reproducible and sufficiently strong band (figure 3.2) was obtained using Optimase™ DNA polymerase (Transgenomic Inc, San Jose, CA, USA) and the conditions shown in table 3.1 and figure 3.3 were used for subsequent analysis. Whilst there is some background smearing and several faint bands, the high concentration in the band of interest gave confidence that the product was suitable for purification and sequencing.

Table 3.1 Volumes and concentrations of reagents used in PCR amplification of DHFR exon 2m (d2m). Total volume = 50 µL.

	Original concentration	Working concentration	Volume (µL)
Buffer (inc, $MgSO_4$)	10x	1x	5.0
dNTP Master Mix	12.5 mmol/L	0.8 mmol/L	3.2
Primer d2m forward	5 µmol/L	0.8 µmol/L	8.0
Primer d2m reverse	5 µmol/L	0.8 µmol/L	8.0
DMSO	100 % $^{V}/_{V}$	2.5 % $^{V}/_{V}$	1.2
DNA	Approximately 100 ng		2.0
Optimase™	2.5 U/µL	2.5 U/50 µL	1.0
Water			21.6

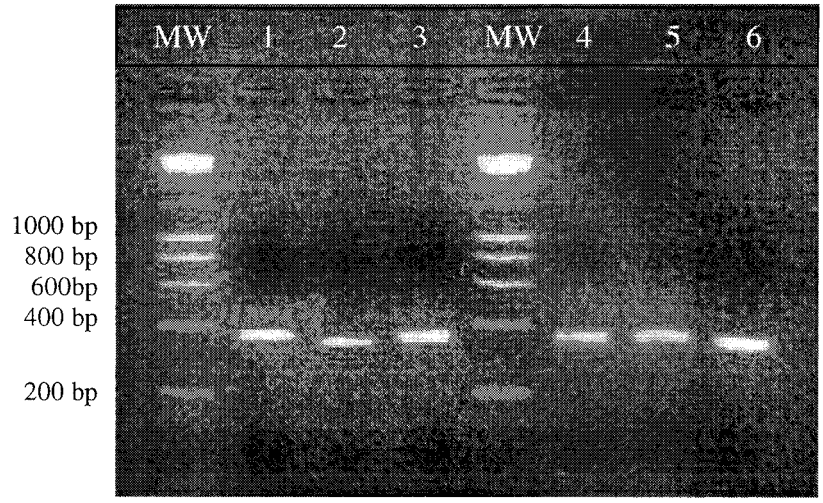
Figure 3.3 PCR conditions for DHFR exon 2m amplification. (Times in minutes)



Sequencing results for NALM6 suggested that there was more interference than that expected from the faint secondary bands, so PCR on alternate sources of DNA was performed. Close scrutiny of the agarose gel (figure 3.4) suggests that the product in some wells may be different sizes and that the intense band obtained for NALM6 could in fact be two bands very close in size. The most likely cause of this would be

an insertion or deletion and would explain the poor sequencing results. Therefore DNA used to prepare the gel in figure 3.4 was purified and sent for sequencing. Sequencing confirmed that different sized products were being amplified and will be discussed in section 3.4.2.

Figure 3.4 Agarose gel showing different migration distances for six samples; 1 = patient L532, 2 = L308, 3 = NALM6, 4 = N4, 5 = L383, 6 = L302



3.1.3 Design of exon 3 primers and optimisation of PCR

After checking for specificity using BLAST searches, the following pair with a predicted annealing temperature of 64°C was selected from those generated by the Primo Pro3.4 software (*http://www.changbioscience.com/primo/primo.html*).

DHFR Exon 3 for	CCC AGC CCT GGA GAA AAC AC
DHFR Exon 3 rev	ACC CAG CTG CCA ATT CTG CC

PCR using Amplitaq Gold produced a single product of the expected 398 bp size (figure 3.5) and subsequent optimisation of temperature and magnesium concentration was performed. The higher temperatures proved to give a more intense band and a magnesium chloride concentration of 2.0 mmol/L was considered optimum. Betaine

at a concentration of 3% gave a marginal improvement (not shown) and therefore was not used. Final concentrations and PCR conditions used for amplification of samples are shown in table 3.2 and figure 3.6.

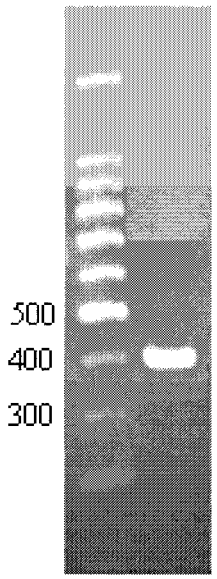
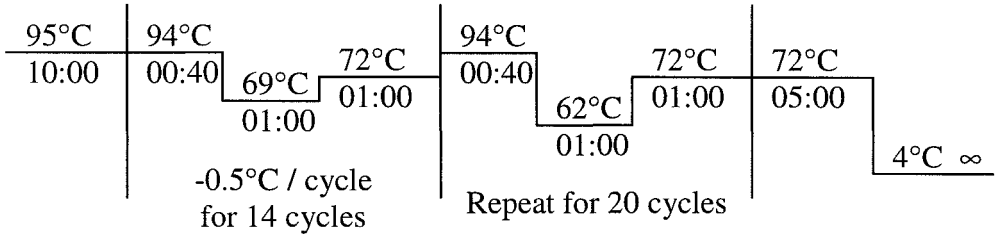


Figure 3.5. Initial PCR showing size of PCR product to be approximately 400bp (expected 398bp) alongside a molecular weight ladder.

Table 3.2 Volumes and concentrations of reagents used in PCR amplification of DHFR exon 3. Total volume = 50 µL.

	Original concentration	Working concentration	Volume (µL)
Buffer	10x	1x	5.0
$MgCl_2$	25 mmol/L	2.0 mmol/L	4.0
dNTP	12.5 mmol/L	100 µmol/L	0.4
Primer d3 forward	5 µmol/L	0.2 µmol/L	2.0
Primer d3 reverse	5 µmol/L	0.2 µmol/L	2.0
DNA	Approximately 100 ng		2.0
AmpliTaq Gold®	5 U/µL	1.25 U/50µL	0.25
Water			19.4
Betaine	10%	3%	15.0

Figure 3.6 PCR conditions for DHFR exon 3 amplification. (Times in minutes)



3.1.4 Design of exon 4 primers and optimisation of PCR

Primer design using several software packages was unable to give a product of optimal size, therefore the smallest at 658 bp was selected for optimisation.

DHFR Exon 4 for	AGA CTC CAC ACA GAC GGT GG
DHFR Exon 4 rev	AGG GTT GGG TCC AGA AAG GG

PCR using standard conditions for Amplitaq Gold produced a single product of approx 650 bp (figure 3.7). A temperature gradient showed that the primers could anneal equally across a relatively wide range of temperature, but 62°C was selected for further work, and that 3% betaine enhanced amplification. Final concentrations and PCR conditions used for amplification of samples are shown in table 3.3 and figure 3.8.

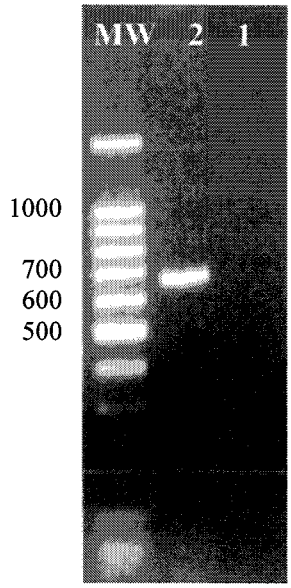
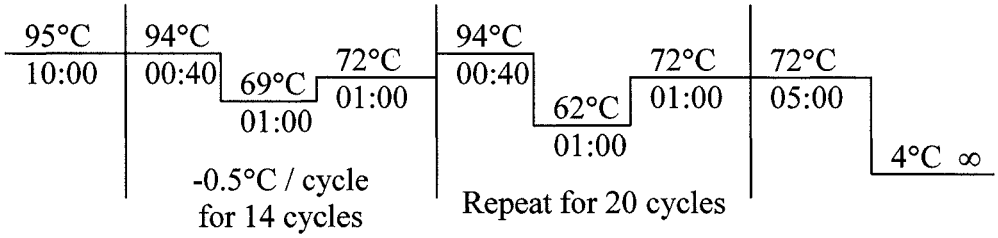


Figure 3.7. Agarose gel showing size of PCR product from initial amplification of primers.
Lane 1 = negative control, 2 = K562.

Table 3.3 Volumes and concentrations of reagents used in PCR amplification of DHFR exon 4.
Total volume = 50 µL.

	Original concentration	Working concentration	Volume (µL)
Buffer	10x	1x	5.0
$MgCl_2$	25 mmol/L	2.5 mmol/L	5.0
dNTP	12.5 mmol/L	100 µmol/L	0.4
Primer d4 forward	5 µmol/L	0.2 µmol/L	2.0
Primer d4 reverse	5 µmol/L	0.2 µmol/L	2.0
DNA	Approximately 100 ng		2.0
AmpliTaq Gold®	5 U/µL	1.25 U/50µL	0.25
Water			18.4
Betaine	10%	3%	15.0

Figure 3.8 PCR conditions for DHFR exon 4 amplification. (Times in minutes)



3.1.5 Design of exon 5 primers and optimisation of PCR

From a number of primer designed the following pair which generated a product of 356 bp in size (see figure 3.9) was considered worthy of further optimisation

DHFR Exon 5 for	CTG ATG TTA AGT GCT TTT TGT TGA
DHFR Exon 5 rev	TTA TAC CTG TTT CTT CCA CTT CCT

Magnesium chloride was required at the standard concentration of 2.5 mmol/L and a temperature of 56°C determined as the optimum. The addition of 2% betaine improved amplification slightly, whilst DMSO had no beneficial effect. Final concentrations and conditions used for amplification of relapse and normal samples are shown in table 3.4 and figure 3.10.

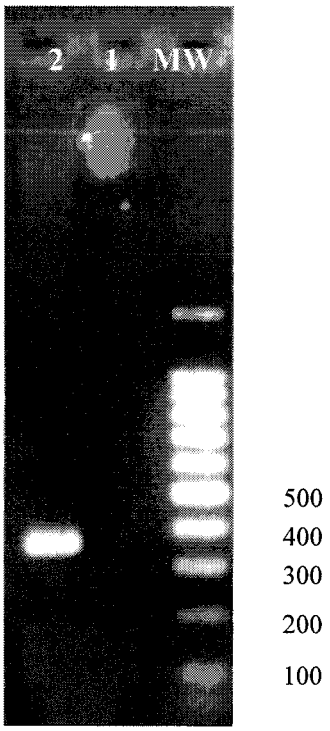
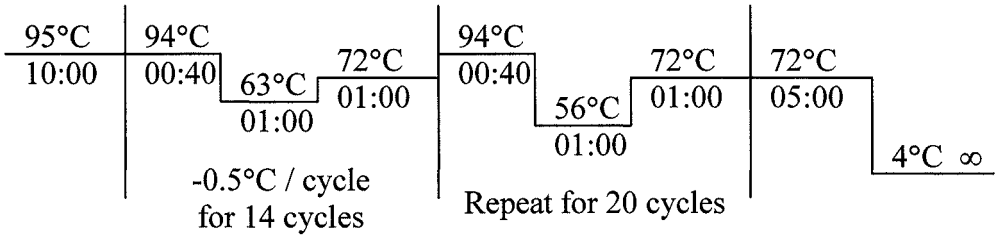


Figure 3.9 Initial PCR of DHFR exon 5 primers showing a product of expected size.
Lane 1 = negative control, 2 = NALM6

Table 3.4 Volumes and concentrations of reagents used in PCR amplification of DHFR exon 5. Total volume = 50 µL.

	Original concentration	Working concentration	Volume (µL)
Buffer	10x	1x	5.0
$MgCl_2$	25 mmol/L	2.5 mmol/L	5.0
dNTP	12.5 mmol/L	100 µmol/L	0.4
Primer d5 forward	5 µmol/L	0.2 µmol/L	2.0
Primer d5 reverse	5 µmol/L	0.2 µmol/L	2.0
DNA	Approximately 100 ng		2.0
AmpliTaq Gold®	5 U/µL	1.25 U/50µL	0.25
Water			23.4
Betaine	10%	2%	10.0

Figure 3.10 PCR conditions for DHFR exon 5 amplification. (Times in minutes)



3.1.6 Design of exon 6 primers and optimisation of PCR

Previously used primer design packages failed to generate a suitable pair due to the reverse primer showing alignment to sequences on chromosomes 1, 13 and 15. Although the chance of amplifying a product from the other chromosomes is unlikely, the primers were discarded at this early stage. Invitrogen OligoPerfect Designer (available at *http://www.invitrogen.com/content.cfm?pageid=1*), generated a single pair with a product size of 351 bp (figure 3.11) and a proposed annealing temperature of 60.1°C. BLAST searches showed there was alignment only with the expected region of chromosome 5 and this pair, (below), was used for further optimisation.

DHFR Exon 6 for	GGC AGC ACC AAG CAT ATT TT
DHFR Exon 6 rev	GCA CCC ATC ATC CTA GCA GT

The optimum temperature was found to be 55°C and the optimum magnesium chloride concentration determined to be 3.0 mmol/L. Three percent betaine gave a slight advantage. Final PCR conditions are shown in table 3.5 and figure 3.12.

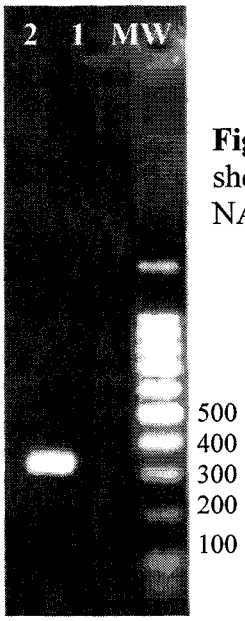
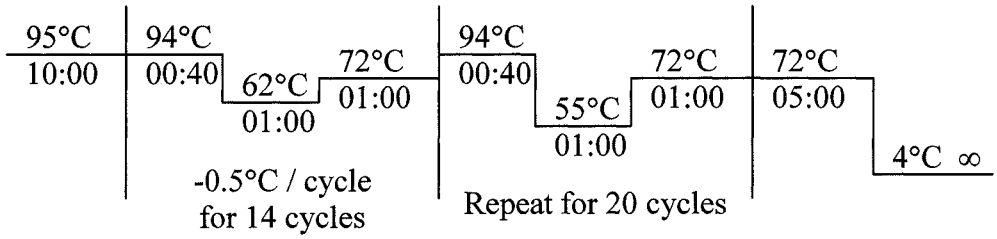


Figure 3.11. Gel showing Initial PCR using of exon 6 primers showing a product of expected size. Lane 1 = negative control, 2 = NALM6

Table 3.5 Volumes and concentrations of reagents used in PCR amplification of DHFR exon 6. Total volume = 50 µL.

	Original concentration	Working concentration	Volume (µL)
Buffer	10x	1x	5.0
$MgCl_2$	25 mmol/L	3.0 mmol/L	5.0
dNTP	12.5 mmol/L	100 µmol/L	0.4
Primer d6 forward	5 µmol/L	0.2 µmol/L	2.0
Primer d6 reverse	5 µmol/L	0.2 µmol/L	2.0
DNA	Approximately 100 ng		2.0
AmpliTaq Gold®	5 U/µL	1.25 U/50µL	0.25
Water			18.4
Betaine	10%	3%	15.0

Figure 3.12 PCR conditions for DHFR exon 6 amplification. (Times in minutes)



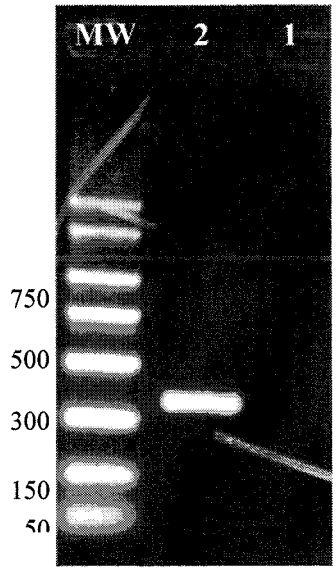
3.1.7 Design of exon 7 primers and optimisation of PCR

By a process of elimination, the following pair was selected as the most specific, with a product size of 351 bp (see figure 3.13) and an annealing temperature of 60.0°C.

DHFR Exon 7 for	CTG AGA ATC AGG GAA GCT GTG
DHFR Exon 7 rev	GAC TCA GTT GGG GTC TTG GA

Temperature optimisation indicated that there was little difference in amplification efficiency across the temperature range (gel not shown) and therefore 60°C was used in subsequent PCR. Likewise, the optimum magnesium chloride concentration was the standard recommended by Applied Biosystems, and neither the addition of betaine nor DMSO made any appreciable difference to amplification efficiency. Therefore standard amplification conditions were used as shown in table 2.3 and figure 2.7

Figure 3.13. Initial PCR of exon 7 primers showing a product of expected size (351 bp).



3.2 CONFIRMATION OF SEQUENCE

Once PCR conditions had been optimised the NALM6 product was purified as described in section 2.5.6. The sequence report was then loaded into Omiga and alignment performed. In each case, alignment confirmed that each sequence containing the relevant exon (shown in appendices 7 – 12).

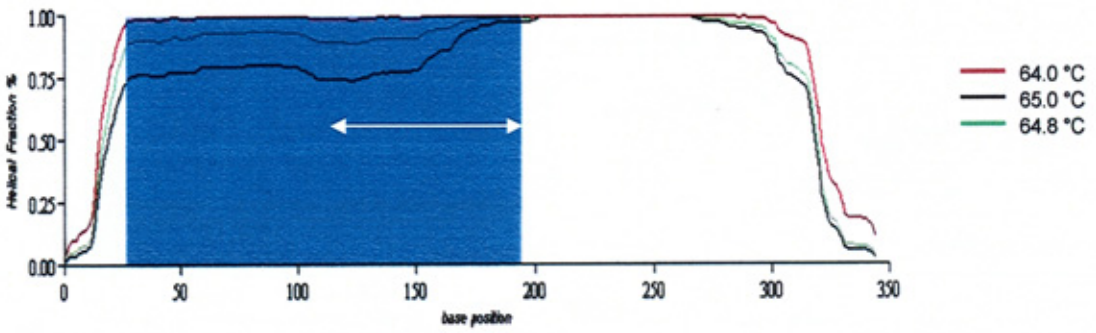
3.3 OPTIMISATION OF WAVE CONDITIONS

With the exception of exon 2, which is discussed later, the PCR product of each exon amplified from NALM6 was used to check DHPLC conditions. The expected wild-type sequence of the product was entered into NAVIGATOR™ and the software generated the optimum temperature for analysis. This was modified depending on the melting profile of the sequence as described in sections 3.3.1-3.3.6. The DHPLC system was prepared for use as described in section 2.5.7.1 and gDNA prepared as described in section 2.5.8

3.3.1 Optimisation of WAVE conditions for analysis of exon 2m PCR products

Analysis of the melt map for this product (figure 3.14) shows that there are principally two domains which are likely to require different temperatures.

Figure 3.14. Melp map of product from DHFR exon 2m primers containing exon 2 (shaded area); coding region shown by arrow.



The exon itself lies between 30 and 195 bases and melts at a lower temperature than the region from 195 – 300 bases. The section between 220 – 260 bases is GC-rich and a predicted temperature of 67°C would be required to give 80% in the helical form. Initial WAVE analysis using three temperatures, 64.8, 65.0 and 67.0°C (figure 3.15) demonstrated that a heteroduplex pattern produced by NALM6 was picked up at the predicted temperature. Subsequent analysis of more samples indicated that analysis at 64.8 or 67°C did not offer any immediate advantage in identifying potential heteroduplexes. The run at 65°C using the conditions shown in table 3.6 was able to differentiate the three patterns and therefore it was decided to run the normal and relapse samples at this single temperature.

Figure 3.15. Chromatograms for CCRF-CEM (homoduplex) on the left and NALM6 (heteroduplex) on the right at 64.8 (red), 65.0 (blue) and 67.0°C (green).

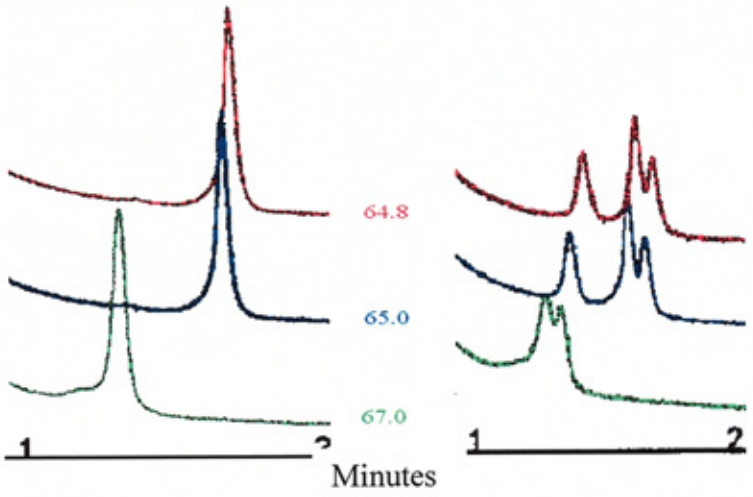


Table 3.6. W AVE run conditions used for analysis of DHFR exon 2 PCR product at 65.0°C. A = 0.1M TEAA, B = 1.2M TEAA + 25% ACN

		65°C	
	Time (mins)	% A	% B
Loading	0	42.5	57.5
Start gradient	0.1	39.5	60.5
Stop gradient	2.1	29.5	70.5
Start clean	2.2	42.5	57.5
Stop clean	2.3	42.5	57.5
Start equilibrate	2.4	42.5	57.5
Stop equilibrate	2.5	42.5	57.5

3.3.2 Optimisation of WAVE conditions for analysis of exon 3 PCR products

As can be seen from the melt map profile in figure 3.16 the coding sequence of interest is that between 110 – 160 bp and lies in a relatively stable region which is 80% helical at 56.5°C. Whilst this is likely to be sufficient to identify heterozygous changes in this region, DHPLC was also carried out at 54.5°C to look for any SNPs that could be used for control purposes or mutations that may have some effect on the neighbouring coding sequence. DHPLC conditions used are shown in table 3.7

Figure 3.16 Melt map generated by NAVIGATOR™ software for the amplified sequence covered by the DHFR exon 3 primers at four temperatures (54.5, 56.5, 56.8 and 59.8°C). Position of exon 3 is indicated by the blue shaded region.

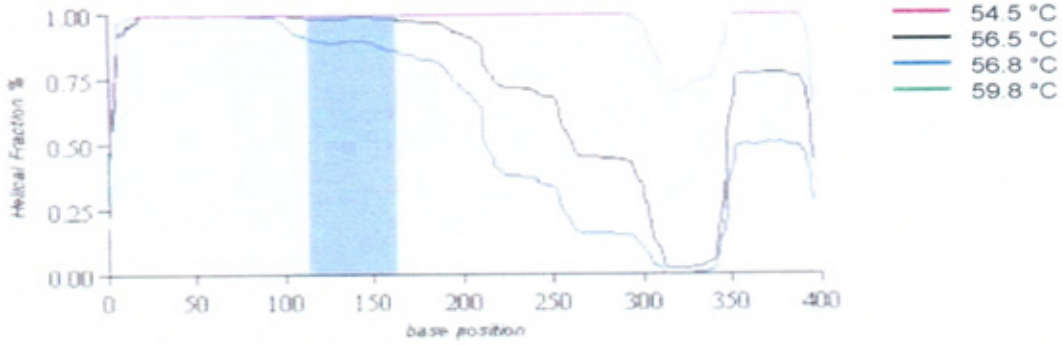


Table 3.7 WAVE run conditions used for analysis of DHFR exon 3 fragments at both 54.5 and 56.5°C. A = 0.1M TEAA, B = 1.2M TEAA + 25% ACN

		54.5°C		56.5°C	
	Time (mins)	% A	% B	% A	% B
Loading	0	42.5	57.5	42.5	57.5
Start gradient	0.1	39.5	60.5	39.5	60.5
Stop gradient	2.1	29.5	70.5	29.5	70.5
Start clean	2.2	42.5	57.5	42.5	57.5
Stop clean	2.3	42.5	57.5	42.5	57.5
Start equilibrate	2.4	42.5	57.5	42.5	57.5
Stop equilibrate	2.5	42.5	57.5	42.5	57.5

3.3.3 Optimisation of WAVE conditions for analysis of exon 4 PCR products

Analysis of the melt map shown in figure 3.17 shows this 658 bp product has several melting domains. Initial optimisation on a small panel of cell-line DNA was tried using 52.9 and 54.0°C. Whilst both were able to demonstrate the presence of heteroduplexes (discussed in section 3.4.4), the pattern was more clearly differentiated at 54°C and therefore was used for screening of the normal and relapse samples. As the column aged, the eluted peaks started to straddle the 'switching surge' at 2.2 minutes (discussed in section 2.5.7.3) and required an offset of –1.0 min and the run conditions shown in table 3.8 to facilitate easier interpretation of the samples.

Figure 3.17 Melt map generated by NAVIGATOR™ software for the amplified sequence covered by the DHFR exon 4 primers at three temperatures (52.9, 53.9 and 54.9°C). Position of exon 4 is indicated by the blue shaded region.

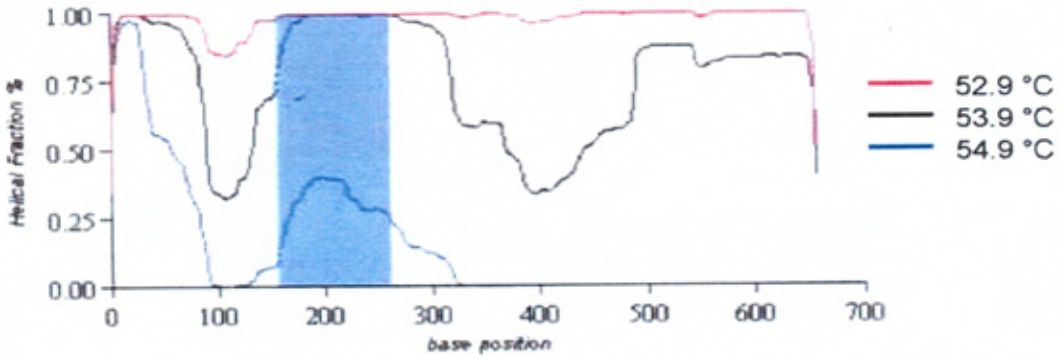


Table 3.8 W AVE run conditions used for analysis of DHFR exon 4 fragments at both 52.9 and 54.0°C. A = 0.1M TEAA, B = 1.2M TEAA + 25% ACN

		52.9°C		Adjusted 54.0°C	
	Time (mins)	% A	% B	% A	% B
Loading	0	40.9	59.1	40.9	59.1
Start gradient	0.1	37.9	62.1	37.9	62.1
Stop gradient	2.1	27.9	72.1	27.9	72.1
Start clean	2.2	40.9	59.1	40.9	59.1
Stop clean	2.3	40.9	59.1	40.9	59.1
Start equilibrate	2.4	40.9	59.1	40.9	59.1
Stop equilibrate	2.5	40.9	59.1	40.9	59.1

3.3.4 Optimisation of WAVE conditions for analysis of exon 5 PCR products

The melt map shown in figure 3.18 shows that the coding region of interest lies across a plateau region. Whilst the coding region is not particularly GC-rich the flanking sequences have lower GC content and this causes the software to calculate a temperature that is average for the whole PCR product, but may not be suitable for the region of interest. Further use of the NAVIGATOR™ software showed that a temperature of 56.2°C would give 80% helical form along the region of interest. Therefore a panel of cell-line DNA was amplified and DHPLC analysis carried out at both 54.7 and 56.2°C. Elution was too fast at the higher temperature and produced a chromatogram with poor resolution. As a result a +0.5 minute adjustment was made and used for the cell-line panel. Based on results from these preliminary experiments, analysis of the normal and relapse samples used only the predicted temperature of 54.7°C using the conditions shown in table 3.9.

Figure 3.18 Melt map generated by NAVIGATOR™ software for the amplified sequence covered by the DHFR exon 5 primers at 53.7, 54.7 and 55.7°C in the top graph; 54.7, 56.2 and 56.3°C in the lower graph. Position of exon 5 is indicated by the blue shaded region in the top graph.

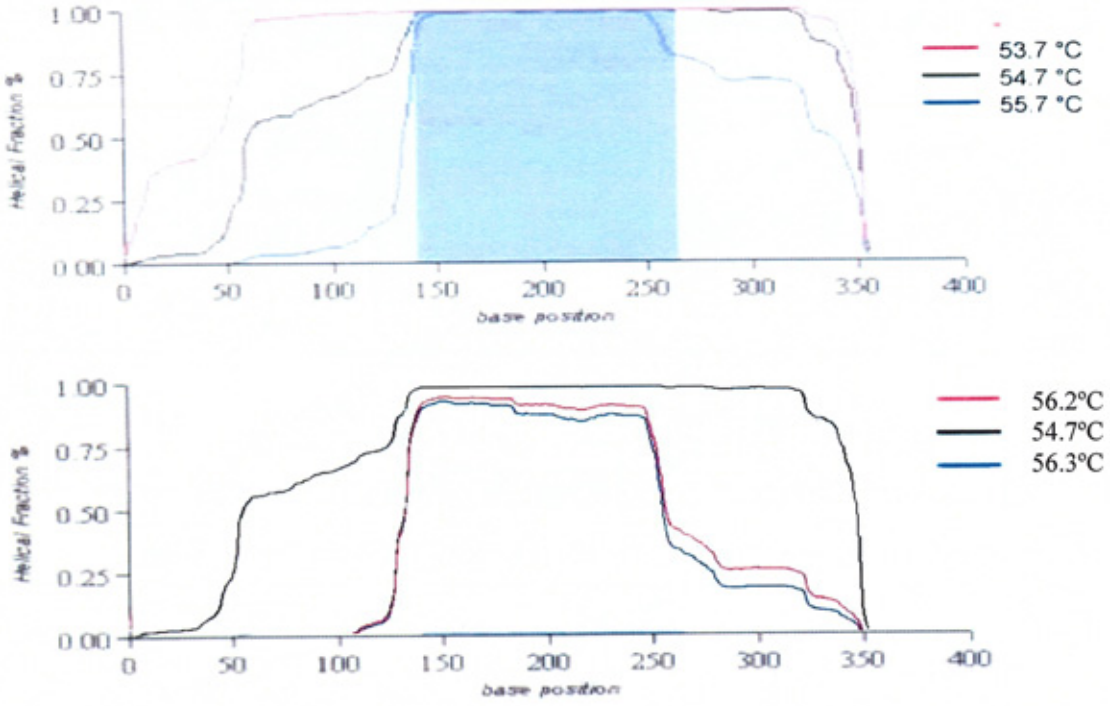


Table 3.9 W AVE run conditions used for analysis of DHFR exon 5 fragments at both 54.7 and 56.2°C. A = 0.1M TEAA, B = 1.2M TEAA + 25% ACN

		54.7°C		Adjusted 56.2°C	
	Time (mins)	% A	% B	% A	% B
Loading	0	44.8	55.2	44.8	55.2
Start gradient	0.1	41.8	58.2	41.8	58.2
Stop gradient	2.1	31.8	68.2	31.8	68.2
Start clean	2.2	44.8	55.2	44.8	55.2
Stop clean	2.3	44.8	55.2	44.8	55.2
Start equilibrate	2.4	44.8	55.2	44.8	55.2
Stop equilibrate	2.5	44.8	55.2	44.8	55.2

3.3.5 Optimisation of WAVE conditions for analysis of exon 6 PCR products

The melt map for the product of exon 6 primers (figure 3.19) shows the coding sequence has a single central melting domain with the 3'-end of the exon having a slightly lower GC content. The predicted 55.2°C temperature proved to be suitable using the chromatographic conditions shown in table 3.10.

Figure 3.19 Melt map generated by NAVIGATOR™ software for the amplified sequence covered by the DHFR exon 6 primers at 55.2, 55.9 and 56.2°C.

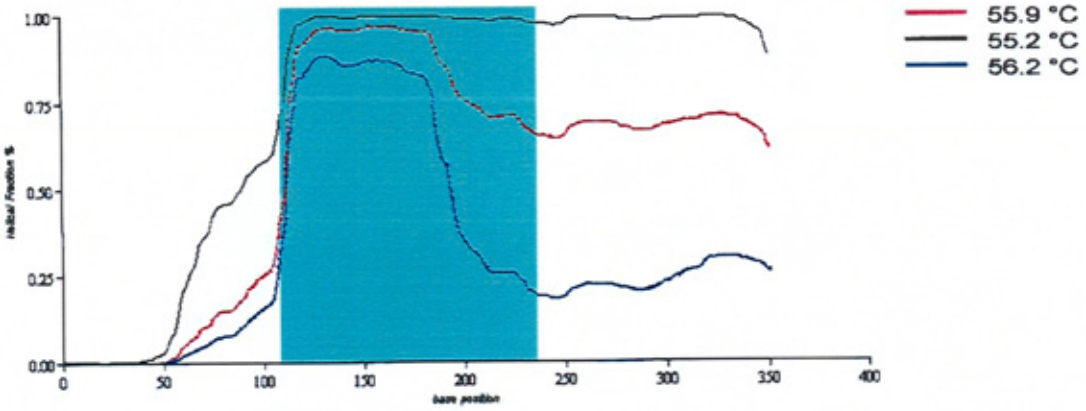


Table 3.10 W AVE run conditions used for analysis of DHFR exon 6 fragments at both 55.2.

		55.2°C	
	Time (mins)	% A	% B
Loading	0	45.0	55.0
Start gradient	0.1	42.0	58.0
Stop gradient	2.1	32.0	68.0
Start clean	2.2	45.0	55.0
Stop clean	2.3	45.0	55.0
Start equilibrate	2.4	45.0	55.0
Stop equilibrate	2.5	45.0	55.0

3.3.6 Optimisation of WAVE conditions for analysis of exon 7 PCR products

There are four melting domains seen on the melt map for the PCR product generated by the exon 7 primers (figure 3.20). The 5'-end of the exon has a higher GC content compared with the rest of the sequence and thus skews the average temperature calculated by NAVIGATOR™. To ensure that the region between 200 and 250 is appropriately screened then 53.6°C was used in addition to the proposed temperature of 54.6°C using the DHPLC parameters shown in table 3.11.

Figure 3.20 Melt map generated by NAVIGATOR™ software for the amplified sequence covered by the DHFR exon 7 primers at 53.6 and 54.6°C.

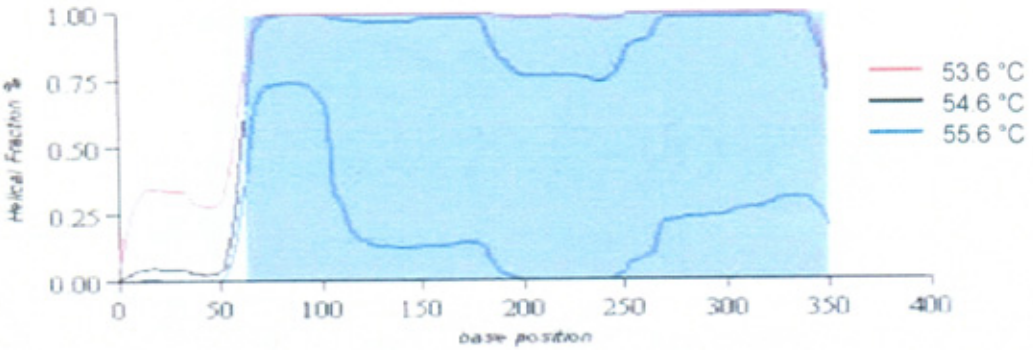


Table 3.11. W AVE run conditions used for analysis of DHFR exon 7 fragments at both 53.6 and 54.6°C.

		53.6°C		54.6°C	
	Time (mins)	% A	% B	% A	% B
Loading	0	45.0	55.0	45.0	55.0
Start gradient	0.1	42.0	58.0	42.0	58.0
Stop gradient	2.1	32.0	68.0	32.0	68.0
Start clean	2.2	45.0	55.0	45.0	55.0
Stop clean	2.3	45.0	55.0	45.0	55.0
Start equilibrate	2.4	45.0	55.0	45.0	55.0
Stop equilibrate	2.5	45.0	55.0	45.0	55.0

3.4 RESULTS

3.4.1 Scoring of WAVE® chromatograms

Performance of the WAVE® was checked as described in section 2.6.1 and with the exception of exon 2, each of the DHFR chromatograms was compared against the homoduplex wild-type pattern generated from NALM6 DNA.

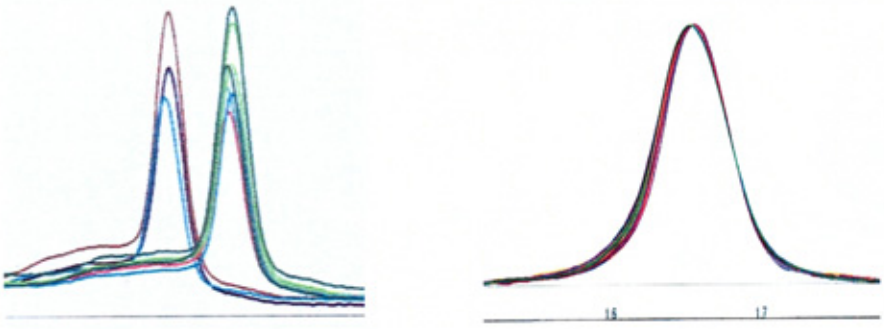
3.4.2 DHFR exon 2

Initial WAVE® runs using NALM6 produced heteroduplex patterns (figure 3.15) that provided further evidence that the poor results encountered on sequencing were the result of a problem at the PCR stage and not at the sequencing or DHPLC stage. When a small panel of alternative DNA were analysed, CCRF-CEM produced a homoduplex pattern and a 'clean' sequence which was used as the wild type for spiking and chromatogram comparisons. The confirmed sequence for CCRF-CEM is shown in appendix 7.2 and will be discussed in section 3.5.1.

3.4.3 DHFR exon 3

All samples, at both temperatures produced homoduplex patterns before and after spiking with NALM6. There was a slight change in retention time between runs (figure 3.21), but using the normalisation function of WAVEMAKER™ all data showed there to be no real difference in the patterns (figure 3.21) and suggests that this was simply the result of the column aging between the runs. Therefore all samples were scored as homozygous wild-type for this exon.

Figure 3.21 (Left) Homoduplex patterns for DHFR exon 3 from two runs showing a shift in retention time. (Right) The same data normalised to show a single homoduplex pattern.



3.4.4 DHFR exon 4

A small panel of cell-line DNA, including K562, HeLa, TK6 and Jurkat were amplified and the chromatograms compared with that of homoduplex NALM6. K562 and Jurkat were the same as NALM6, but HeLa and TK6 produced heteroduplex patterns. At both 52.9 and 54.0°C heteroduplex patterns were observed as clearly different from homoduplex NALM6 (figure 3.22). This difference was further enhanced when the traces were normalised (figure 3.23). The heteroduplex patterns were not very well differentiated at 52.9°C and did not appear to add any advantage to the screening, so this temperature was not used for the screening of the normal and relapse samples.

Figure 3.22 Homoduplex patterns from NALM6 (left) and heteroduplex from TK6 at both 52.9 (red) and 54.0°C (green).

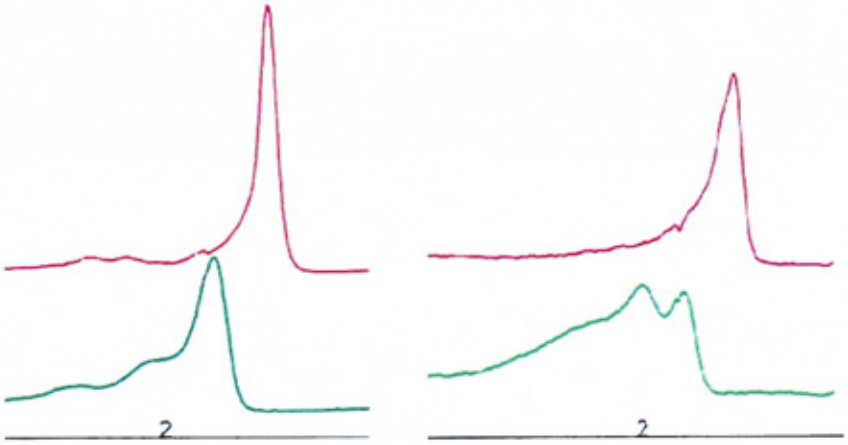
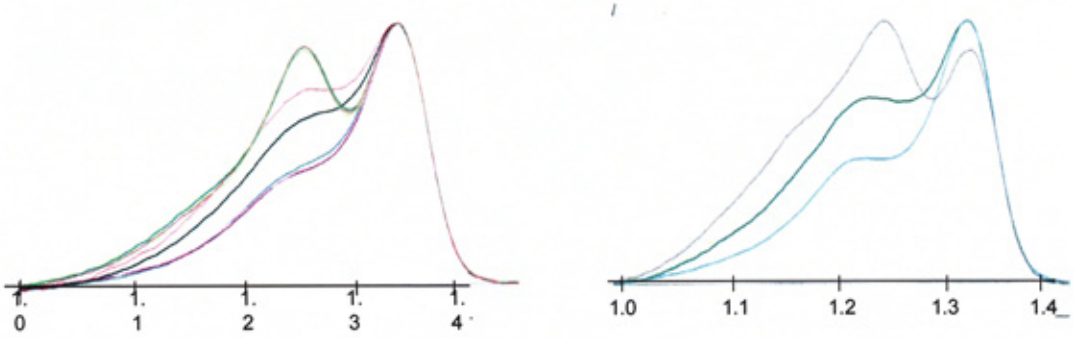


Figure 3.23 (Left) Normalised traces to show the differences between the homoduplex (single peak at 1.31 minutes), the two peak heteroduplex with the additional peak at 1.24 minutes, and the intermediary pattern with the peak at 1.24 minutes presenting as an emphasised pre-peak shoulder.
(Right) Normalised traces of spiked samples showing that similar homoduplex and heteroduplex patterns were obtained.



The majority of samples were scored as homoduplex (see appendix 9.4), but some gave heteroduplex patterns. Samples 189 and 359 gave the 2 peaked trace seen for TK6, whilst 126 and 318 produced a pre-peak shoulder which is clearly seen in the normalised graphs shown in figure 3.23, but was not obvious on the standard chromatogram (not shown). When spiked some of the normal and relapse samples started to produce a heteroduplex pattern, indicating homozygous genotypes different to that of NALM6 (figure 3.32). The chromatograms were similar to those shown in figure 3.22, but to confirm the genotype sequencing is required. Samples sent for

sequencing failed. As these were cell-lines made up of one clone then it was unlikely that the method was picking up a mixed DNA which could have been present in relapse sample DNA Amplification, purification and sequencing will be performed when 'expanded' material become available.

3.4.5 DHFR exon 5

NALM 6 gave the expected homoduplex pattern at both 54.7 and 56.2°C as shown in figure 3.24. The small panel of cell-line DNA analysed as a preliminary test of the DHPLC conditions produced heteroduplex patterns shown in figure 3.25. In all cases the heteroduplex pattern was more clearly seen at 54.7°C and therefore only this temperature was used for analysis of the normal and relapse samples.

Figure 3.24 Chromatogram of single peaks for NALM6 exon 5 at 54.7 and 56.2°C.

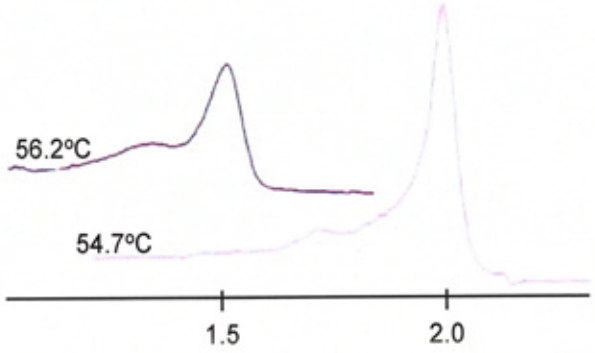
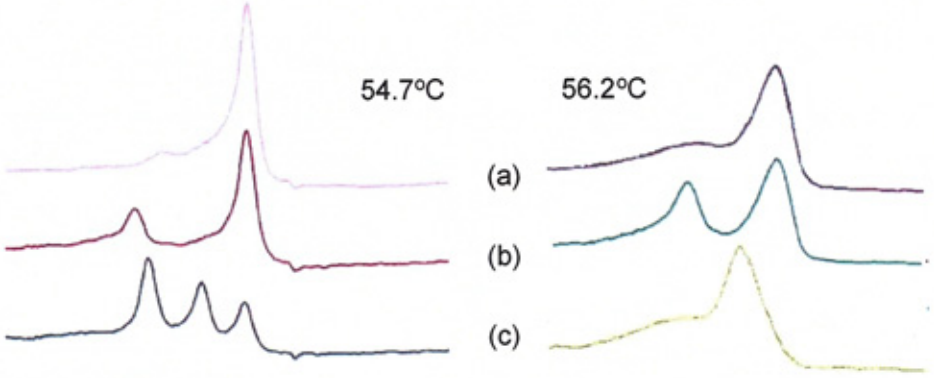
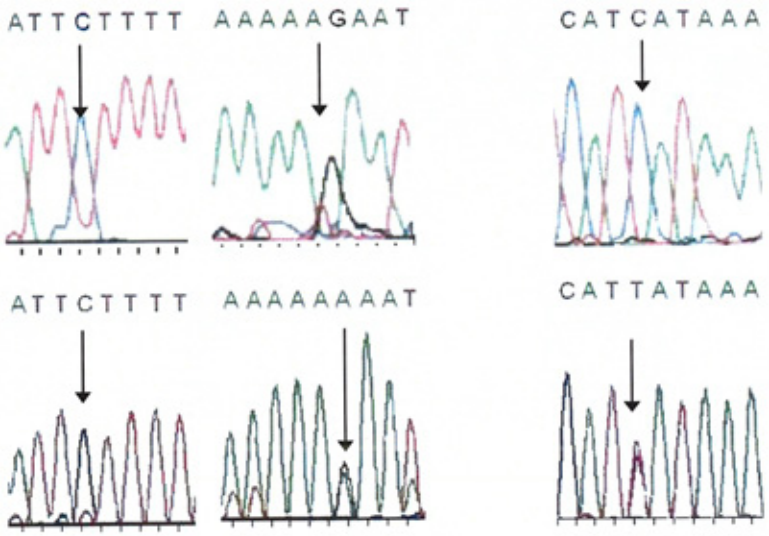


Figure 3.25 Chromatogram showing homoduplex (a) NALM6 and two heteroduplex patterns (b) K562, (c) Jurkat, at both 54.7 and 56.2°C for DHFR exon 5.



The results from sequencing of K562 and Jurkat were aligned with NALM6 to determine the most likely cause of the heteroduplexes. Figure 3.26 shows portions of the forward and reverse genome upstream of the coding sequence that did not align perfectly. The only difference noted for K562 occurred on the reverse sequence with both A and G present in equal amounts at the position indicated by the arrow in figure 3.35 (intron 5-6 base position 11362 and shown on the sequence in appendix 10.1). However this heterozygous result was not evident on the forward sequence.

Figure 3.26 (Left) Sequence upstream of DHFR exon 5. Arrow indicates base position 11362 of intron 5-6. Top row shows forward and reverse sequences for NALM6; bottom row shows the forward and reverse sequences for K562. (Right) The reverse sequence for NALM6 on the top and Jurkat on the bottom. The arrow indicates base position 11279 of intron 5-6.



The only difference noted on the Jurkat sequence was upstream of the coding sequence (intron 4-5 base position 11279 shown in appendix 10). Only the reverse sequence is shown in figure 3.35 because the base position was too close to the forward primer and failed to give a clear result.

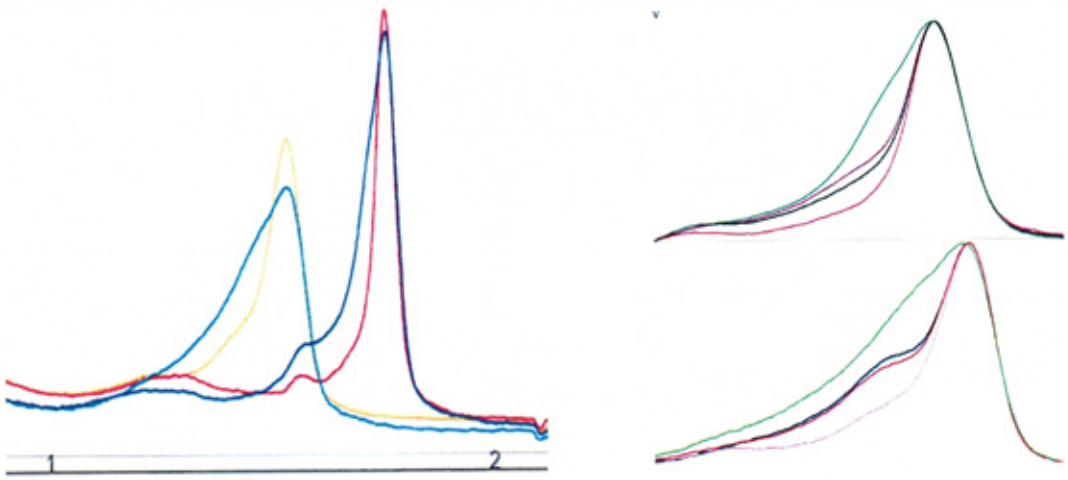
3.4.6 DHFR exon 6

All samples produced homoduplex patterns before and after spiking with NALM6 and suggest that there are unlikely to be any SNPs or mutations in this section of chromosome 5. However, without a positive control to verify that the DHPLC conditions were optimal, then an absolute decision cannot be made.

3.4.7 DHFR exon 7

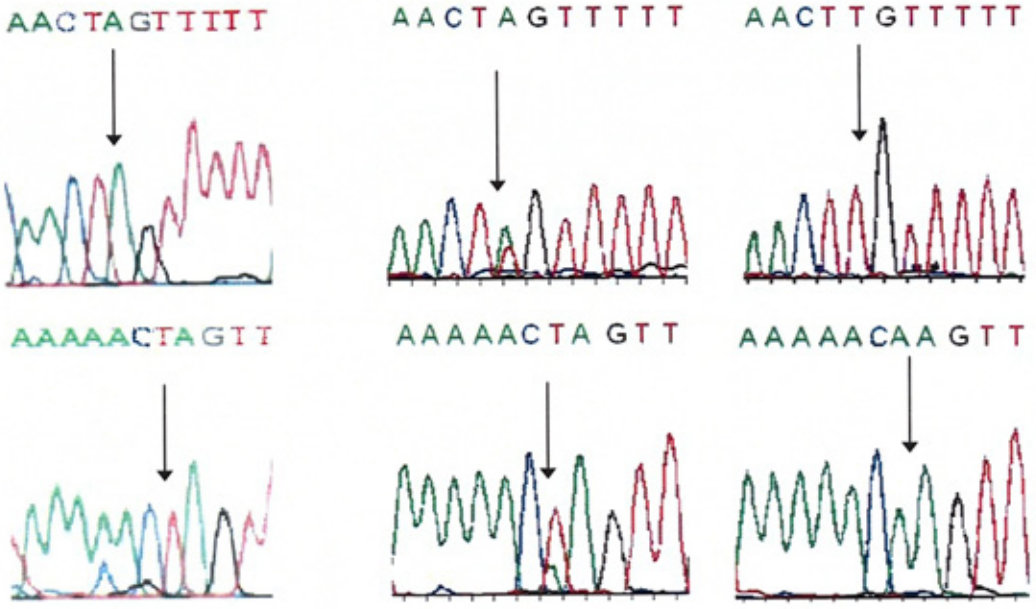
The initial DHPLC run of NALM6 PCR product gave the expected homoduplex pattern at both 53.6 and 54.6°C and was followed up by analysis of DNA from a selection of cell-lines. At 53.6°C K562 did not appear to be much different from NALM6 and would probably have been scored as homoduplex (figure 3.27). But the broader band was exaggerated at 54.6°C and suggested an underlying earlier secondary peak. When the peaks at both temperatures were normalised then the differences are much clearer and K562 was scored as a heteroduplex (figures 3.27).

Figure 3.27. (left) NALM6 homoduplex patterns, red at 53.6°C, yellow at 54.6°C; overlaid with heteroduplex pattern of K562, dark blue at 53.6 and light blue at 54.6°C. (right) Normalised chromatograms at 53.6°C (top) and 54.6°C (bottom) showing heteroduplex pattern of K562 (green) is clearly differentiated from other samples.



Use of the normalisation function allowed samples to be scored as heteroduplex (appendix 12.4). When spiked with NALM6, two of the normal samples gave a heteroduplex previously not evident, suggesting that they had a homozygous genotype different from the wild type. These two samples and K562 were sequenced to determine what the difference or differences were. Alignment against NALM6 highlighted a base pair in the 3'UTR which was heterozygous in K562 A/T and homozygous T/T in N4 (figure 3.28). This would account for the heteroduplex patterns before and after spiking with NALM6 which is homozygous A/A. The sequence data for the other normal sample was poor and did not allow any comparison with the wild type.

Figure 3.28. Sequence in the 3'UTR down stream from exon 7 which is likely to account for the heteroduplex patterns. Forward sequence shown on the top, reverse on the bottom for NALM6 (left), K562 (centre) and N2/4 (right). Position of interest shown by an arrow.



3.5 DISCUSSION

Exons 3 and 6 gave homoduplex patterns for all samples before and after spiking indicating that these exons did not contain any SNPs or mutations.

3.5.1 DHFR exon 2

The coding sequence of exon 2 has a GC content of 62%, but the regions either side used for the design of primers have much higher content, 70% for that upstream and 82% downstream. This may partly explain the difficulty experienced in designing and optimising primers, also made difficult by the relatively common 19 base pair deletion in this region.

Of the 50 normal and relapse samples with sufficiently strong signal to score; 68% gave the three peak heteroduplex pattern originally seen for NALM6, 10% gave a two peak heteroduplex and 22% were scored as homoduplexes. When the 3 peak heteroduplex pattern was sequenced (figure 3.29) it was apparent that this sequence contained the site of a previously reported deletion (Johnson *et al.,* 2004; 2005). In this study 78% of samples had the -19/-19 homozygous or WT/-19 heterozygous genotypes which compares well with the 80% in Johnson's (2005) group of pregnant women. Applying the Fisher exact probability test to determine the statistical difference between small populations (*http://faculaty.vassar.edu/lowry/fisher.html*) then a 2-tailed p-value of 0.0166 indicates that the difference between presentation and relapse samples is significant. None of the relapse samples had the WT/WT (+19/+19) genotype. The significance of this must be considered with the findings of Parle-McDermott *et al.,* (2007), who reported a slight increase in mRNA. They point out that the increase was not statistically significant, but if a larger cohort could be studied then it should be possible to confirm if there is a real increase or whether the

conflicting findings of Xu *et al* (2007), in which they propose a loss of a transcription binding site, are more correct.

Figure 3.29 Forward sequence showing -19/-19 homozygous (top), and +19/-19 heterozygous sequences.

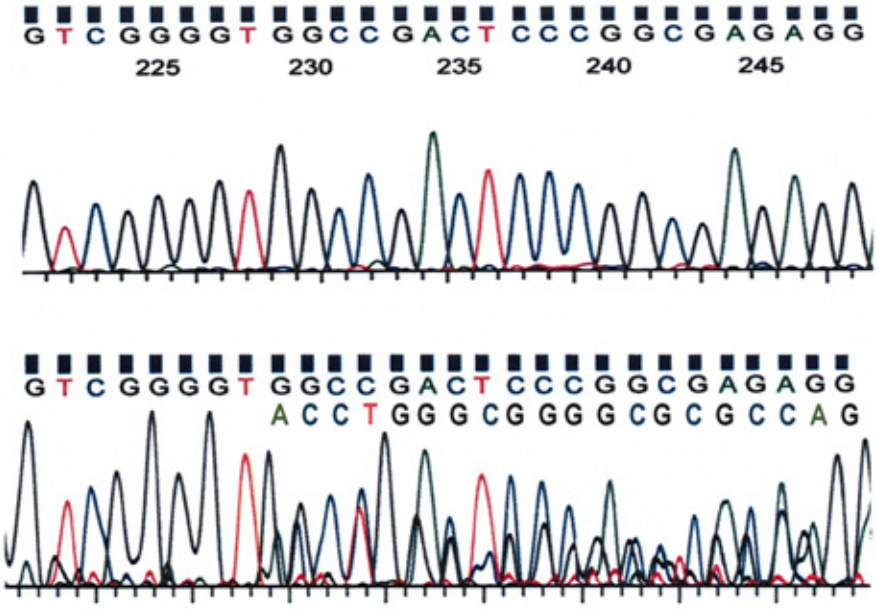
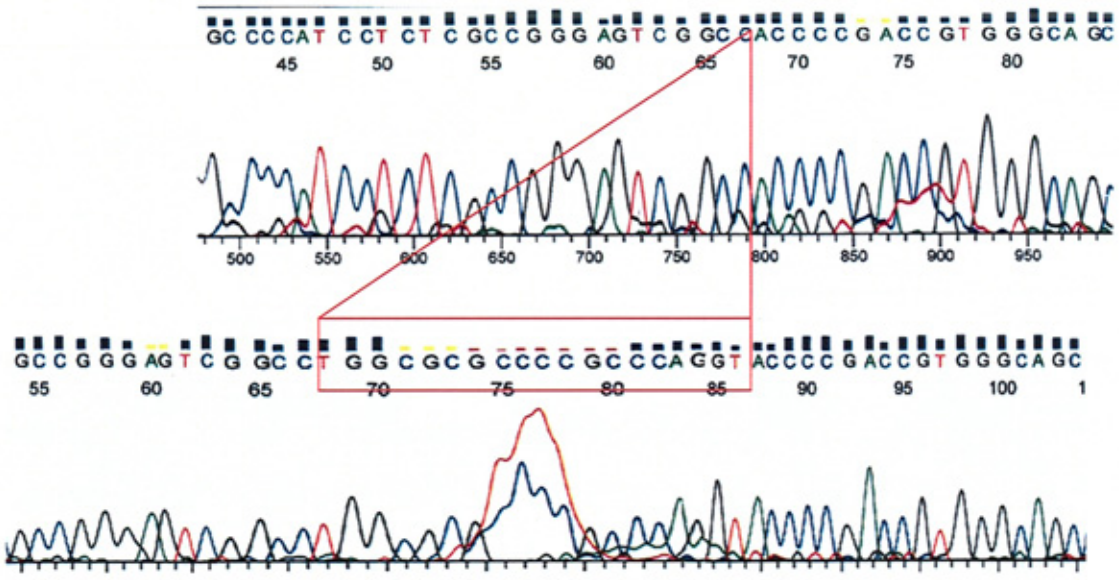


Figure 3.30 Reverse sequence showing -19/-19 homozygous (top), and +19/+19 (bottom). The +19 bases are highlighted in the box and linked to the insertion point in the top sequence.



The original report of this deletion by Johnson *et al* (2004) proposed that the loss of these 19 bases simply alters the distance between transcription factor binding sites and the start codon. In either case, it is clear that this polymorphism does have a role to play in transcription of the gene and folate metabolism (Stanisławska-Sachadyn *et al.*, 2008; Scionti *et al.*, 2008). Reducing the amount of active DHFR may increase sensitivity of the cell to MTX.

When sequenced the 2 peak heteroduplex pattern seen in sample L382 appears to be the result of a T/C heterozygous state at position -25, upstream of the start codon, but this was only observed on the reverse sequence for this sample (figure 3.31). Similarly, the 2 peak pattern seen for relapse sample 6652 (figure 3.32) had a A/G heterozygous base call. These two options could not be satisfactorily discriminated and without sufficient evidence the genotyping for the rest of the unknown samples was not made from the chromatograms. Further work is required to repeat and confirm these sequences as the cause of the heteroduplex patterns.

Figure 3.31 Forward (top) and reverse (bottom) sequences from CCRF-CEM (left) and for sample L382 (right). Arrow indicating base position -25 from the start codon,

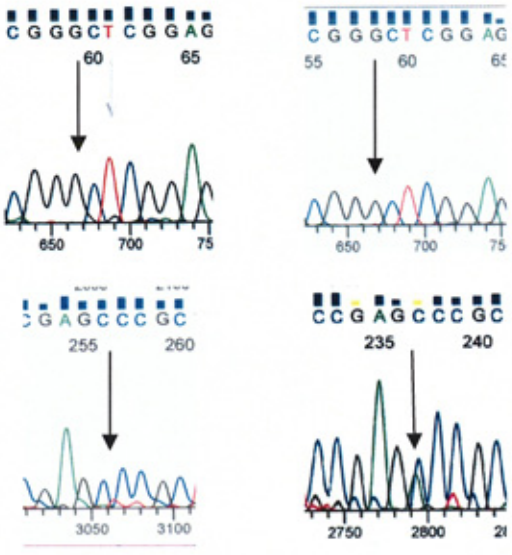
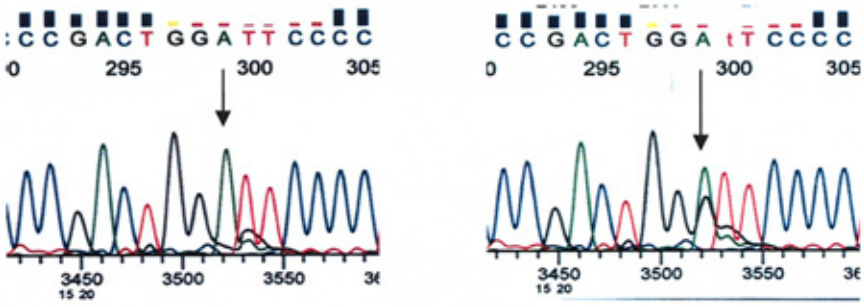


Figure 3.32 Forward sequences from homoduplex CCRF-CEM (left) and for heteroduplex sample 6652 (right) showing A/G at base position.



3.5.2 DHFR exon 4

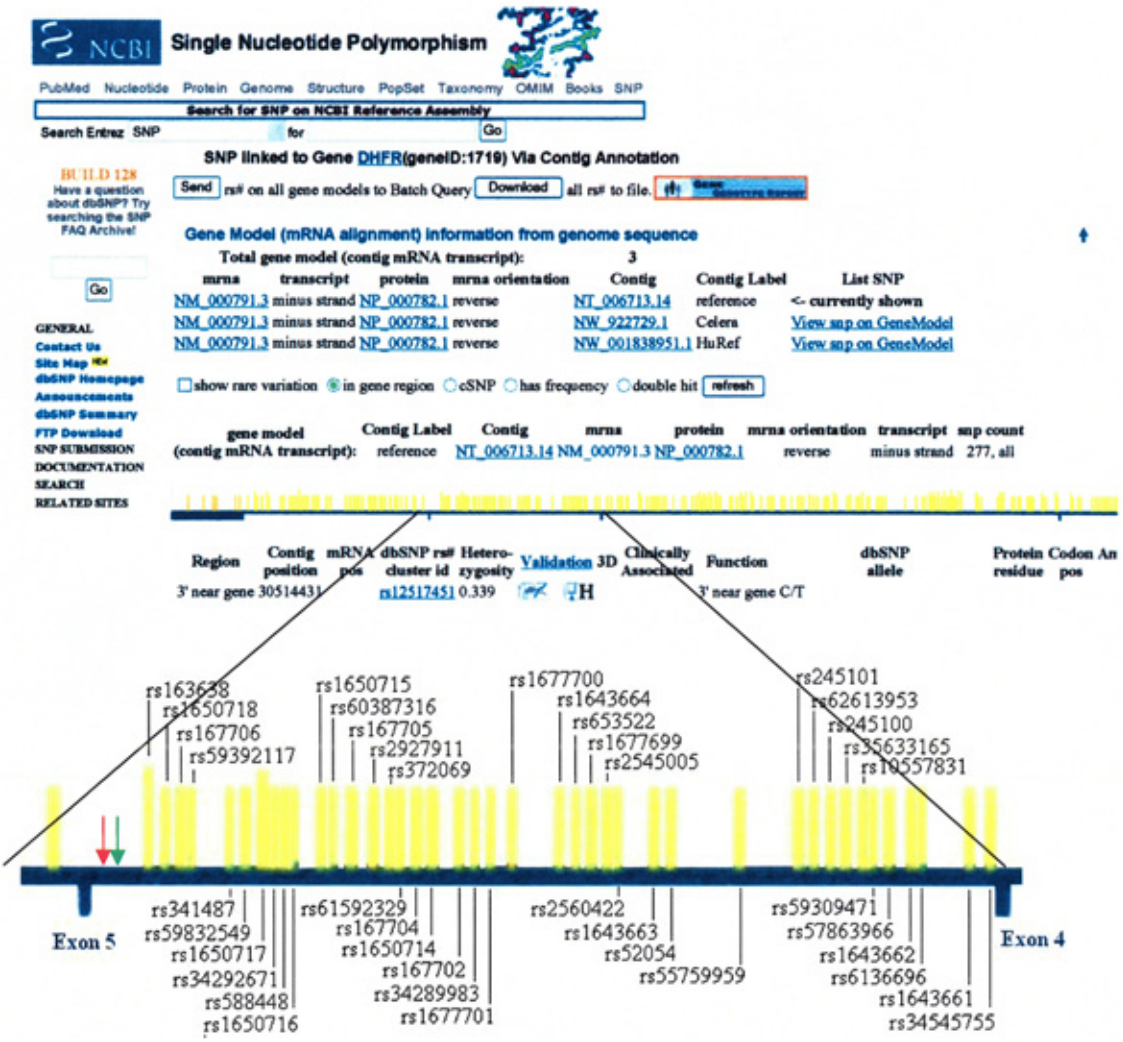
The NCBI SNP database was searched on the 30/4/2008 for a polymorphism that might account for the heteroduplex patterns seen but nothing of interest was obtained. On the 5/7/2008 the Cancer Genome Anatomy Project web site was interrogated and this proposed a SNP at position 164 (A/G) of exon 4 recorded as candidate SNP ID 2422154 (this does not correspond to rs2422154 on the NCBI SNP database). NALM6 appears to have the wild type A/A genotype at this position. Sequencing of samples with heteroduplex patterns will confirm is this is the cause of the heteroduplexes.

3.5.3 DHFR exon 5

The two base positions in the amplified sequence which probably account for the heteroduplex patterns seen by DHPLC were not positively identified as any of the reported polymorphisms reported on the NCBI SNP database (*http://www.ncbi.nlm.nih.gov/sites/entrez*) last searched 24/3/2008. (shown on figure 3.33). Both locations are upstream of the coding sequence, sufficiently distant from the splice sites and not restricted to either normal or relapse groups to suggest there is any contribution to MTX resistance.

Using Chi-squared, no statistical difference (p = 0.63) was observed between the frequency of heteroduplex patterns in the normal group compared to the relapse group (see appendix 10.4 and 10.5).

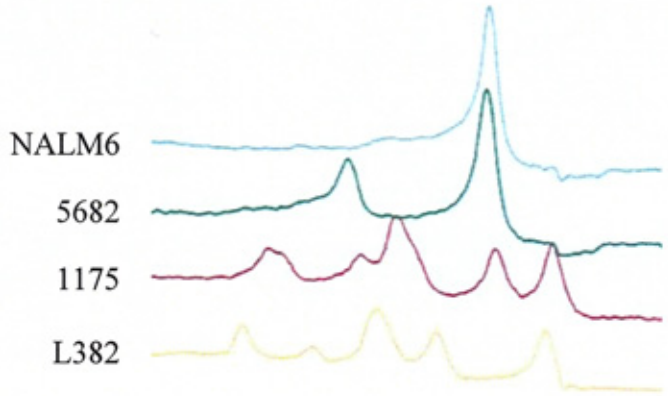
Figure 3.33 Chromosome 5 segment adapted from NCBI SNP database showing locations of known SNPs in intron 4-5. Locations of sequence differences found in K562 (base 11362) and Jurkat (base 11279) are shown by red and green arrows respectively.



Four of the relapse samples with heteroduplex patterns were amplified, purified and sequenced to determine if the difference was within the coding region. Samples 8678 (L5) and 5862 (L169) gave the two peak pattern first seen for K562. Sequence

alignment also confirmed the similarity to the genotype of K562, ie., heterozygous A/G on the reverse sequence at intron 4-5 base number 11362. Each of the two peak heteroduplex patterns were subsequently scored as this A/G heterozygote. The sequence data from samples 1533 (L345) and 7096 (L37) gave good forward results but nothing was noted that could account for the four peak patterns seen in figure 3.34. The reverse sequence was unreadable due to a secondary set of peaks at each base position which made it impossible to interpret each base correctly. This could be due to an insertion or deletion in one allele and requires repeat analysis once sufficient DNA becomes available on these patients.

Figure 3.34 Showing the single peak homoduplex pattern for NALM6, the double peak pattern seen for 5862, and the multiple peak heteroduplex patterns seen for samples 1175 and L382.



3.5.4 DHFR exon 7

The screening method was able to pick up the sequence differences shown in figure 3.31, but none of these reside in the coding sequence and therefore are unlikely to play any role in MTX resistance.

3.6 CONCLUSION

Considering the importance of this enzyme to normal cell metabolism there was a lack of published research into the gene using gDNA. The design and optimisation of primers suitable for this purpose has now been done. Used here they have confirmed that mutations or SNPs in the coding sequence are unlikely and that MTX resistance involving DHFR *in vivo* is most likely at the level of transcription. Whilst there is still more work to do to confirm the genomic difference identified by DHPLC in the PCR product from exon 4 primers, a greater focus should be applied to understand the significance of the 19 base pair deletion in the 5'-UTR of exon 2. In particular, increasing the sample size to verify that the difference observed between the presentation and relapse groups is real, and hopefully go on to explain the possible mechanism.

4 GENES INVOLVED IN FOLATE AND METHOTREXATE METABOLISM

It is well recognised that sub-typing of neoplastic disease can facilitate more focused and therefore effective treatment (Yeoh *et al.*, 2002; Mosquera-Caro *et al.*, 2003; Ross *et al.*, 2003; Wright *et al.*, 2003; Dobbin *et al.*, 2005; Carroll *et al.*, 2006; Hoffman *et al.*, 2008). Several methods have already been described in section 1.2 which are used to find out as much as possible about the malignant cells in leukaemia and this has contributed significantly in achieving the long-term survival success in childhood ALL. A relatively new tool, adding to this information, is the microarray.

4.1 GENOME-WIDE MICROARRAYS

One advantage the microarray brings to molecular biology is that it facilitates multiple genes to be examined in a single event (Dalma-Weiszhausz and Murphy, 2002; Chua *et al.*, 2006; Mullighan *et al.*, 2007). This has allowed patterns in gene expression and copy number to be seen which had previously not even been considered (Miller *et al.*, 2004; Butcher, 2007; Sorich *et al.*, 2008). Sequential analysis using a microarray following administration of a drug has provided invaluable information to the pharmaceutical industry on how the cell responds in a holistic manner and has elucidated links in metabolic pathways, many of which were unknown (Gullans, 2006; Sorich *et al.*, 2008). The importance of this is that it could help identify potential side-effects of new drugs before going to clinical trials and indicate new categories of drugs that may be of use in alternative situations. Sequential analysis during the cell cycle has also provided information on when and which genes are switched on at different points and provide potential novel targets for therapy. With the improved inter-laboratory precision described by Dobbin *et al* (2005) and bioinformatics tools that allow computational analysis across databases at different

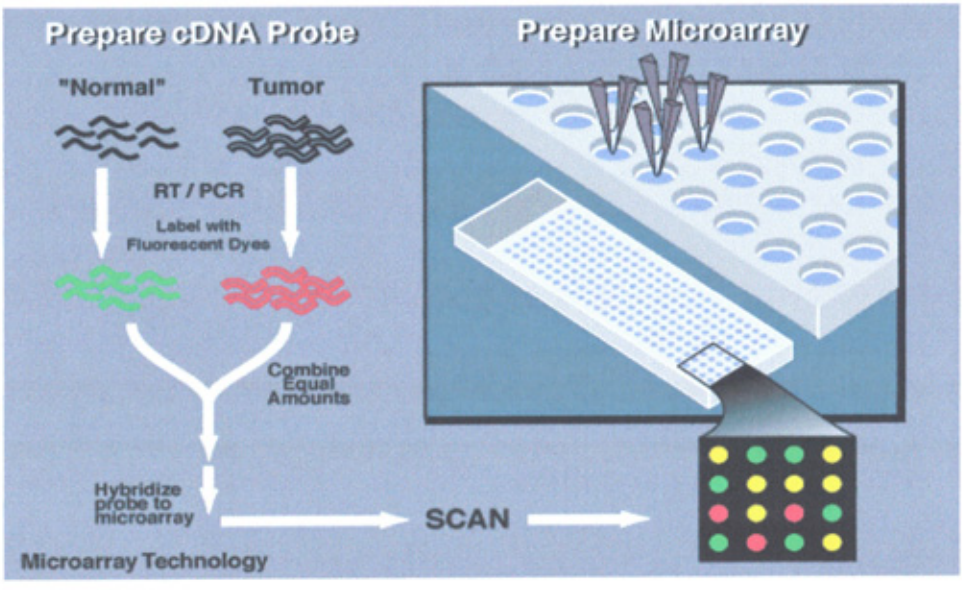
levels (Hanauer *et al.*, 2007), it should be possible to diagnose and screen for cancers and other diseases with greater sensitivity and specificity by their unique pattern of gene expression (Choudhuri, 2004; Dunphy, 2006; Mi *et al.*, 2007).

In chapter 1, a number of enzymes involved in folate (1-carbon) metabolism were introduced that can potentially affect MTX metabolism or be influenced by the anti-folate properties of MTX, directly or indirectly. Use of microarray technology to provide data about these folate-associated proteins may provide patterns that suggest mechanisms for somatic or acquired MTX resistance.

4.2 PRINCIPLES OF MICROARRAY APPLICATION

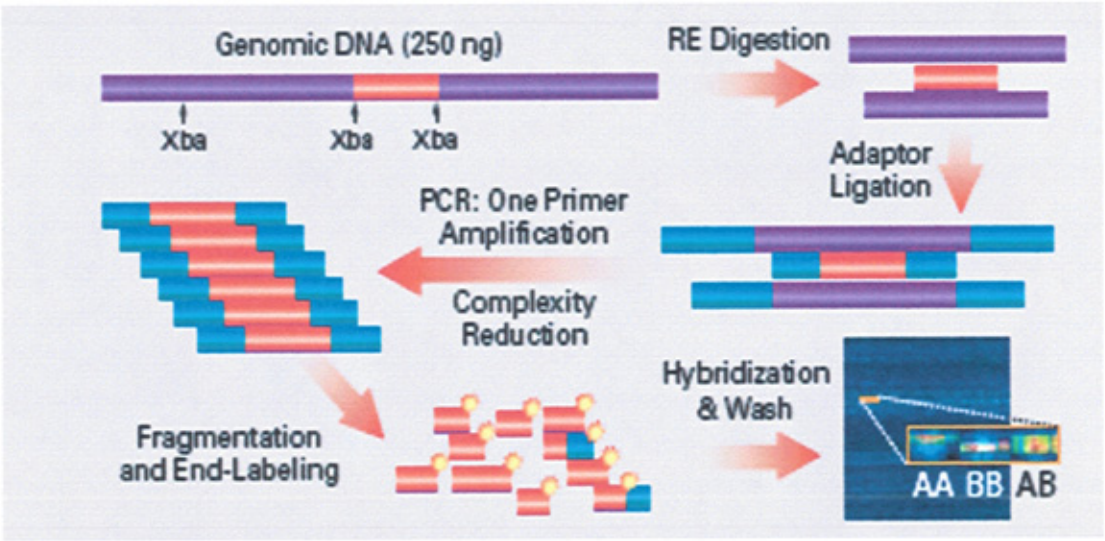
There are two main formats for analysis of genes using microarrays (van't Veer *et al.*, 2002; Ferl *et al.*, 2003; Barrett and Kawasaki, 2003; Carroll *et al.*, 2006; Dunphy, 2006). The first is the cDNA microarray in which cDNA from a library of sequences is immobilised onto the glass or nylon slide (Schena *et al.*, 1995). The second is the oligonucleotide microarray in which specifically designed oligonucleotides are built onto the slide.

Figure 4.1. Basic principle of microarray technology. (Taken from NHGRI)



If the array is being used to study gene expression (ie., mRNA levels), then the mRNA in the sample is used as the template for reverse transcription PCR, during which a fluorescent label is incorporated into the amplified product. These fluorescent products then bind to the matched sequence on the microarray and provide a signal (figure 4.1). Alternatively, when screening for specific genotypes or copy number the gDNA is digested with a restriction endonuclease to leave sticky ends on the gDNA fragments. Adapters are ligated to the sticky ends and the sample then undergoes PCR with primers designed to the adapters (figure 4.2).

Figure 4.2 Diagram showing formation of fluorescent fragments of gDNA to be hybridised to Affymetrix GeneChip® (Anon, 2004)



This non-specifically amplifies the whole of the genome and each of the fragments is subsequently tagged with a red fluor (Cy5). The same is done with a normal sample, but this time the fragments are labelled with a green fluor (Cy3). The two are then mixed and applied to the microarray (figure 4.1). Competitive hybridisation is then allowed to take place and excess DNA washed off. The microarray is scanned by a laser and the fluorescent signal recorded. If there is more normal DNA (ie., loss of copy number) then the spot will fluoresce green; equal amounts of normal and

unknown produce a yellow signal; if the target has increased copy number in the unknown then the spot will fluoresce red (note that different coloured fluors can be used). If the immobilised cDNA is not recognised by any labelled fragment the spot will not fluoresce (Choudhuri, 2004).

Oligonucleotide microarrays require prior knowledge of the sequence of interest to design specific oligonucleotide sequence. These can be synthesised *in vitro* and the 25-mers sprayed on to the slide as small dots. As in primer design, the oligonucleotide sequence must not have any secondary structure or the capacity to anneal with itself, if it is to work correctly. The technique developed by Affymetrix for the GeneChip® (Affymetrix Inc, Santa Clara, CA) uses oligonucleotide sequences 25 bases in length that have been synthesised *in situ* on the glass slide (chip) using photolithographic techniques to build up the required sequence a base at a time (Fodor *et al.*, 1991). Robotic technology has allowed >50,000 spots each measuring 20-200 μm in diameter to be placed on a single 25 mm by 75 mm microscope slide (Barrett and Kawasaki, 2003; Meaburn *et al.*, 2006), each spot containing a different oligonucleotide ligand. *In situ* oligonucleotide synthesis has allowed 500,000 spots to be generated on a 1.28 x 1.28 cm slide (Barrett and Kawasaki, 2003). This increased number of ligands provides greater resolution in examination of the genome. In the 100k SNP GeneChip® just over 100,000 SNPs have been identified from the Perlegen Sciences SNP database which are distributed across, and represent, the human genome (Anon, 2004).

4.3 NORMALISATION AND ANALYSIS OF DATA

In the comparison of data from different chips the biggest problem is the analytical variability of the technique caused by differences in the quality of the DNA, dot

formation on the array, surface characteristics of the slide, differences in binding affinity of labelled sequences and fluorimeter voltage fluctuations. Therefore a method to normalise the data to determine a true signal from background variation is required (Park *et al.*, 2003). Normalisation is the setting of a threshold at which the intensity of the signal sufficiently changes from an equimolar (yellow) signal to either of the dominant forms (red or green signal) and thus represent a true change and not just background variation. Techniques for normalisation have been reviewed by Park *et al* (2003), Cui and Churchill (2003) and more recently by Chua *et al* (2006).

4.4 USE OF MICROARRAYS TO IDENTIFY COPY NUMBER AND LOH

Understanding mechanisms that explain how cells transform into malignant clones was enhanced by reports of loss of heterozygosity (LOH) (Irving *et al.*, 2005). Many genes have alterations in them that never present phenotypically, mainly because the alternative allele is expressed. This difference is clearly understood to be heterozygosity, but if a secondary event 'knocks out' the normal allele there is a loss of heterozygosity and the cell can become malignant. LOH can result from partial or complete loss of a chromosome carrying the gene, breakages and/or translocations (Zheng *et al.*, 2005; Ryder, 2007) or more subtly due to uniparental disomy (Engel, 1998). Uniparental disomy or 'imprinting' is an epigenetic effect that selects for expression of the maternal or paternal allele and enhances the female or male characteristics. This has been found to complicate inheritance studies and cause LOH (Engel, 1998). It becomes a problem in heterozygous disorders, when the normal allele is silenced by this mechanism leaving the defective allele to be expressed, or no allele to be expressed.

On the 50K SNP microarray, fifty thousand SNPs have been selected from across the genome which show the greatest variation in the general population. The probability of one being homozygous reflects the genotype frequency, but is relatively small. When two homozygous SNPs occur next to each other this is less likely to occur by random chance, three even less likely, etc. Therefore the probability of LOH value increases the more homozygous sequential SNPs there are in a region.

Bioinformatics packages available for the interrogation of microarray data facilitate analysis of copy number and LOH (Zhou *et al.*, 2005; Hu *et al.*, 2006). Data available from patients at ALL presentation and relapse were analysed here, to estimate copy number and concurrent probability of LOH in regions in the human genome containing genes involved in folate metabolism that may add to our understanding of relapse.

4.5 METHODS

Genomic DNA from 77 presentation childhood ALL and 20 relapse patients was sent to Affymetrix UK Ltd (High Wycombe) for analysis on their 50 K SNP microarray and compared with a Caucasian normal data set. The raw data were made available via the Newcastle University website for use by the research team at the Northern Institute of Cancer Research. The data are arranged by L number but anonymised to all but the database manager to protect patient confidentiality and ensure unbiased analysis of the data set. Within the raw data is a confidence value that the reading is different to that of the reference normal population for each patient sample at a particular location along the genome (see section 4.3).

The locations of genes involved in folate metabolism and likely candidates for involvement in the development of MTX resistance were identified and the SNP closest to this location selected as the point of interest (shown in table 4.1). To increase the confidence in the interpretation and develop cumulative occurrence scoring for LOH, the data for surrounding sequences was also selected as a subset for analysis and copied to an Excel file from the main database (thanks to Marian Case for help in the preparation of the data subset).

Table 4.1 Showing name and location of key genes involved in folate metabolism used to select a subset of 50 K microarray data.

Protein full name (abbreviation)	**Chromosome** (location)	**ENSEMBL No**	**50 K SNP identifier**
Methylenetetrahdyrofolate reductase (MTHFR)	**1** 11768367-11788702	ENSG00000188985	SNP A-1721350
Methionine synthase (MS)	**1** 235025277-235133904	ENSG00000116984	SNP A-1643875
Aminoimidazole carboxamide ribonucleotide transformylase (AICAR or AICART)	**2** 215885036-215922724	ENSG00000138363	SNP A-1701154
Methionine synthase reductase (MTRR)	**5** 7922217-7954237	ENSG00000124275	SNP A-1691055
Betaine methyltransferase (BMT or BHMT)	**5** 78443358-78463864	ENSG00000145692	SNP A-1749904
Dihydrofolate reductase (DHFR)	**5** 79957800-79986556	ENSG00000188985	SNP A-1673169
Gammaglutamyl hydrolase (GGH) also known as **Folylpolyglutamyl hydrolase** (FPGH)	**8** 64090197-64113965	ENSG00000137563	SNP A-1741370
Folylpolyglutamyl synthetase (FPGS)	**9** 129596697-129616427	ENSG00000136877	SNP A-1703348
Serine hydroxymethyl transferase (SHMT)	**17** 18171920-18207581	ENSG00000176974	SNP A-1659463
Thymidylate synthase (TS)	**18** 647619-663492	ENSG00000176890	SNP A-1731420
Glycinamide ribonucleotide transformylase (GART)	**21** 33836286-33798139	ENSG00000159131	SNP A-1668065
Cystathionine β-synthase (CBS)	**21** 43346374-43369541	ENSG00000160200	SNP A-1687757
Reduced folate carrier (RFC)	**21** 45737914-45786779	ENSG00000173638	SNP A-1754481

4.5.1 Statistical analysis of data

To determine if there is a statistical difference in copy number at regions of the genome containing key folate genes; a t-test to determine if the means of the two sets of data came from the same population was first considered. This is a parametric statistical tool that is only of value if the data is normally distributed. Therefore the values for both presentation and relapse groups were plotted as a histogram to determine if the data in both groups was normally distributed (examples shown in Figures 4.3-4.7).

Figure 4.3. Example of histogram showing almost Gaussian distribution of copy number significance for presentation and relapse samples (SNP-A1661146).

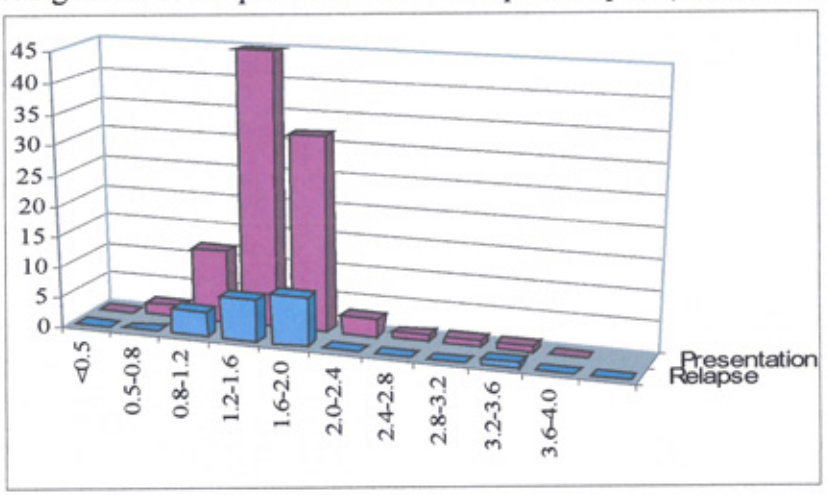


Figure 4.4. Example of histogram showing non-Gaussian distribution of copy number significance for presentation and relapse samples (SNP-A1711211).

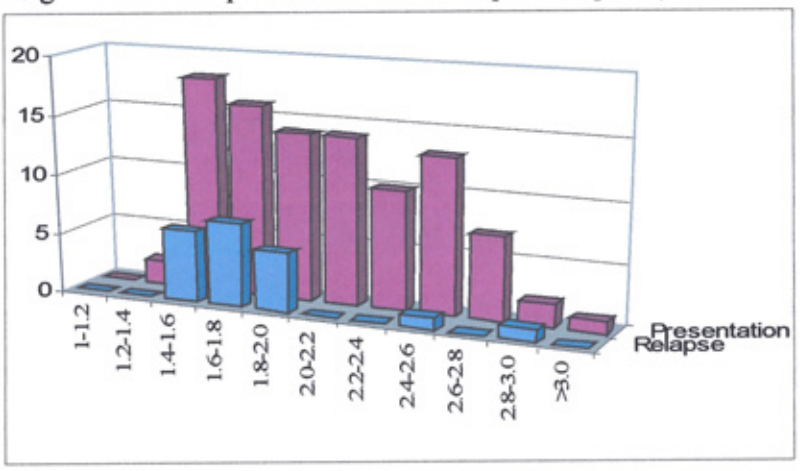


Figure 4.5. Example of histogram showing non-Gaussian distribution of copy number significance for presentation and relapse samples (SNP-A1681341).

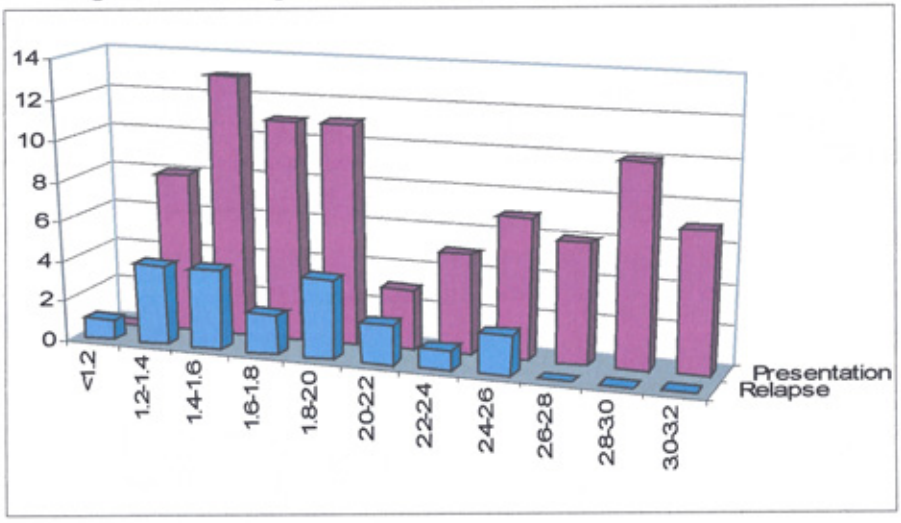


Figure 4.6. Example of histogram showing non-Gaussian distribution of copy number significance for presentation and relapse samples (SNP-A1711211).

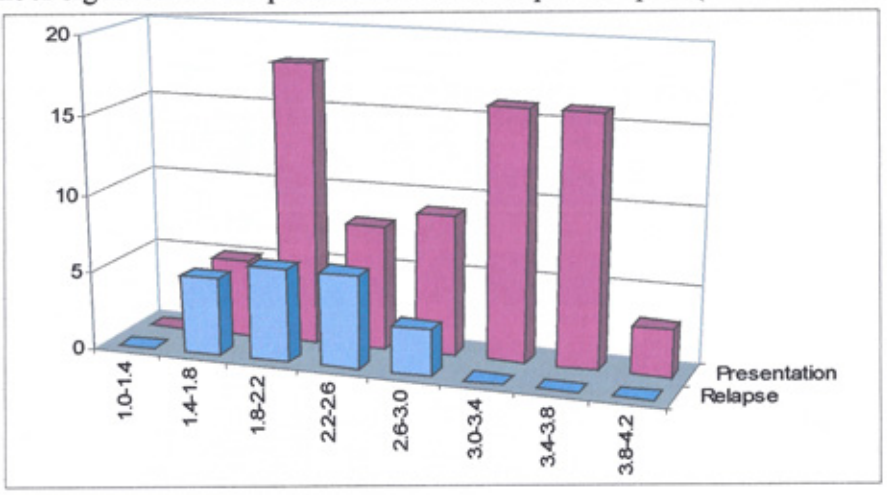
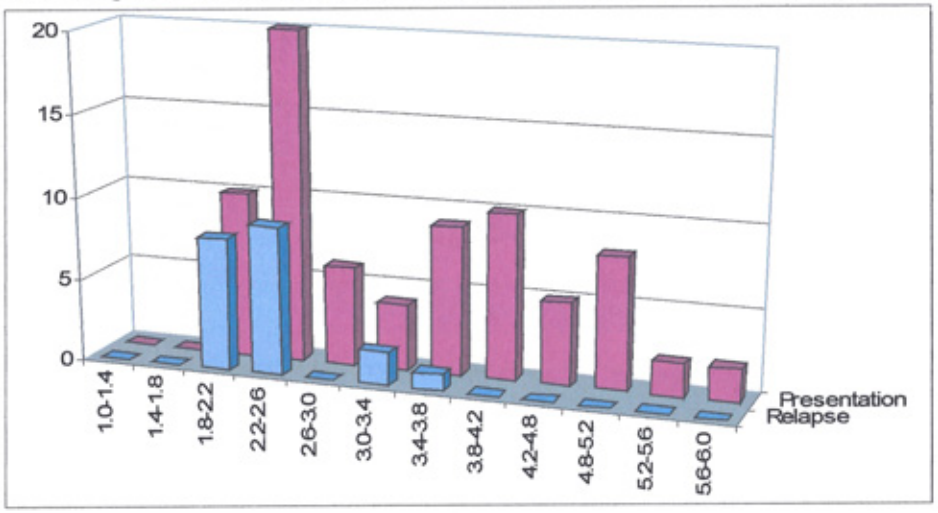


Figure 4.7. Example of histogram showing non-Gaussian distribution of copy number significance for presentation and relapse samples (SNP-A1703348).



After preparing a small number of representative histograms for data from across the genome (figures 4.3 – 4.7) it was soon apparent that the data were not normally distributed, and whilst Chui and Churchill (2003) suggest a log transformation as one solution, it was decided that with the relapse group only having a sample size of 20, nonparametric analysis would be more appropriate.

The Mann-Whitney non-parametric statistic for comparison of population medians (Swinscow, 1983; Hart, 2001; Avery, 2004; Hawkins, 2005) was used to determine if there was a statistical difference in either estimated copy number between presentation and relapse groups and again for probability score of LOH. As the data set only has five paired presentation-relapse samples, then this statistic is preferred over the Wilcoxon Rank test which uses paired data. The Mann-Whitney test assumes a 'null hypothesis', (ie., that there is no difference between the medians of the two populations), and then assigns rankings to the data. Sums of ranked values for each population are used in the calculations and the results compared with values from tables of Rank Sum at a particular confidence limit and defined degrees of freedom (n_1 – 1 against n_2-1). Both upper and lower limits are calculated as no assumption is made about whether the relapse median is higher or lower than the median of the presentation population (ie., 2-tailed). If the calculated values of U_1 or U_2 lie outside the range obtained from the tables then the hypothesis can be rejected at that level of significance. The further the calculated values are from the median of the presentation samples the higher the confidence in rejecting the hypothesis (ie., the smaller the p-value). The freeware statistical package available from (*http://elegans.swmed.edu/~leon/stats/utest.cgi*) was used to perform the calculations.

4.6 RESULTS AND DISCUSSION

In the identification of the closest SNP to the gene of interest, it became apparent that for the RFC gene, the SNP in one direction gave a different result than the SNP in the other direction and validated the reason for looking at extended regions up and downstream of the gene of interest to see if there is any influence on the promoter regions.

Mann-Whitney statistical analysis for the comparison of medians from both relapse and presentation samples provided p-values to determine the statistical significance of any difference. p-values of greater than 0.05 were not considered significant, p-values between 0.01 and 0.05 were significant at the 5% level, p-values between 0.001 and 0.01 were considered significant at the 1% level, p-values <0.001 were considered highly significant at the 0.1% level (ie., reject the 'null hypothesis' with an error rate of only 1:1,000).

The p-values for difference were plotted for each SNP position across the relevant part of the chromosome to graphically demonstrate if the difference was more significant across specific regions and/or in particular the gene of interest. In figures 4.8 to 4.22 the green stripe on graph (a) indicates the location of the gene in relation to the SNP position. From these graphs it was relatively easy to identify those genes with a statically significant differences between the presentation and relapse samples for either estimated copy number (blue line) or probability of LOH (purple line). However, the p-values at each location were derived from the presentation and relapse groups in totality, and this may mask some interesting information about specific samples or sub-population clusters within each group. Therefore, scatter plots were prepared using the SNP data set closest to the gene of interest for both estimated copy

number and probability of LOH. If the LOH data included a very high value, then a second scatter plot is shown using a log scale to expand the lower range and facilitate easier analysis of the data. In graph (a) the small arrow (↓) indicates the SNP position used for the scatterplots. In each of the scatterplots the connecting lines indicate paired samples (pair 83, 173, 181, 479, 554). Boxes show the median and 5 and 95% percentiles.

4.6.1 METHYLENETETRAHYDROFOLATE REDUCTASE (MTHFR)

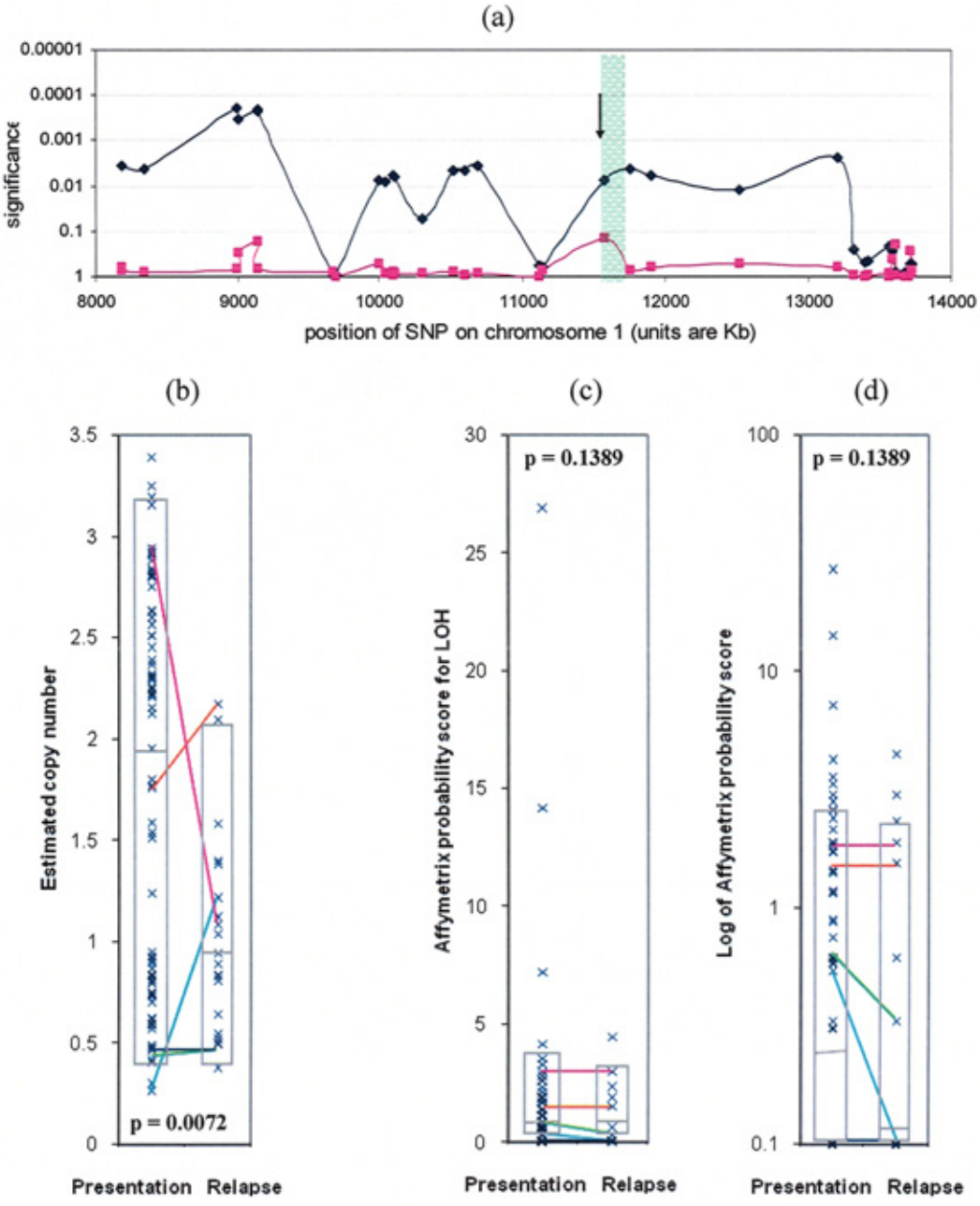
Using the SNP location on the short arm of chromosome 1 closest to the MTHFR gene (see figure 4.8a), the p-value of 0.0072 for estimated copy number is highly significant and from figure 4.8b it can be seen that this is the result of a bimodal distribution of presentation samples in which a group with values between two and three are distinct from presentation samples with values between 0.5 – 1.0 and the relapse group. Such low values did not fit the expected pattern or those obtained for other genes where the estimated copy number was supported by cytogenetic findings. Cadiux *et al.,* (2006), suggest a reduction in copy number in a region of chromosome 1, which includes the MTHFR gene, has a protective role in maintaining normal enzyme activity against the effects of hypomethylation.

Thirty-seven of the seventy presentation samples had a reading above the 95% percentile for the relapse group and formed a distinct population within the presentation group. Looking at neighbouring SNPs in either direction, then this drop in estimated copy number seems to be restricted to this one SNP position (table 4.2).

Table 4.2 Average estimated copy number at other SNP positions surrounding SNP A-1721350 on chromosome 1.

SNP identification	Location	Average estimated copy number
A-1647019	11115517	3.58
A-1652145	11145703	3.58
A-1721350	11565924	1.60
A-1647681	11753710	2.26
A-1703614	11896676	2.68

Figure 4.8 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 1 including the MTHFR gene. (b)(c) and (d) Scatterplots using data from SNP A-1721350



Therefore an alternative theory is that, normalisation of the signal at this point is using an inappropriate reference and the Affymetrix signal could be recalibrated to push the population with values around 2.5 above 3 and indicate that they have extra copy number. Zhou *et al.,* (2004) provide evidence for such a theory when they reported increased copy number of one allele along the q-arm of chromosome 1 in oral malignancy which then exhibits a LOH, but there are no such changes reported in the p-arm of patients with ALL (Harrison, 2001). The t(1;14)(p33;q11) translocation involving chromosome 1 only occurs in 3% of T-cell ALL and therefore is not likely to account for the 53% of presentation samples with the higher copy number. A smaller ~100 kb translocation referred to as the *TALd* occurs in 30% of T-cell ALL (Pui *et a.l*, 1994) and is a more likely candidate, except that this is several megabases from the MTHFR gene.

The findings here warrant further investigation because none of the relapse samples have values greater than 2.5. It would appear that having a higher copy number contributes to a better outcome. Alternatively, if there is a true loss of this gene, then amplification and DNA methylation studies are required to determine if protection against hypomethylation (Cadieux *et al.*, 2006) is the most likely explanation.

The importance of this enzyme, at a point in metabolism where folates can be directed toward remethylation of HCys, or alternatively into thymidine synthesis, is well recognised. The increased copy number and subsequent higher enzyme activity (presumed) could lead to a deficiency of folates for thymidine synthesis. Skibola *et al* (1999) proposed that this increases susceptibility for ALL in adults. The antifolate action of MTX is also to reduce thymidine synthesis and therefore a higher copy number of MTHFR could be expected to enhance the chemotherapeutic effect. This

theory could explain the lack of relapse patients with a high copy number, but this is a small group of patients and more work will be required to support the theory. It will also be necessary to identify if the subgroup have a specific genotype for the known MTHFR polymorphisms (Wiemels *et al.,* 2001; Urano *et al.,* 2003) which will also influence enzyme activity.

Whilst SNPs in this gene are associated with increased risk of cancer (Sharp *et* al., 2002; Yi *et* al., 2002; Quinliven *et* al., 2005), Skibola *et al* (2002) suggested that there was little evidence that SNPs or mutations in MTHFR have any effect on MTX sensitivity. The effect that Urano *et al* reported in the same year, when using MTX in the treatment of rheumatoid arthritis, is probably due to the much lower doses of MTX used.

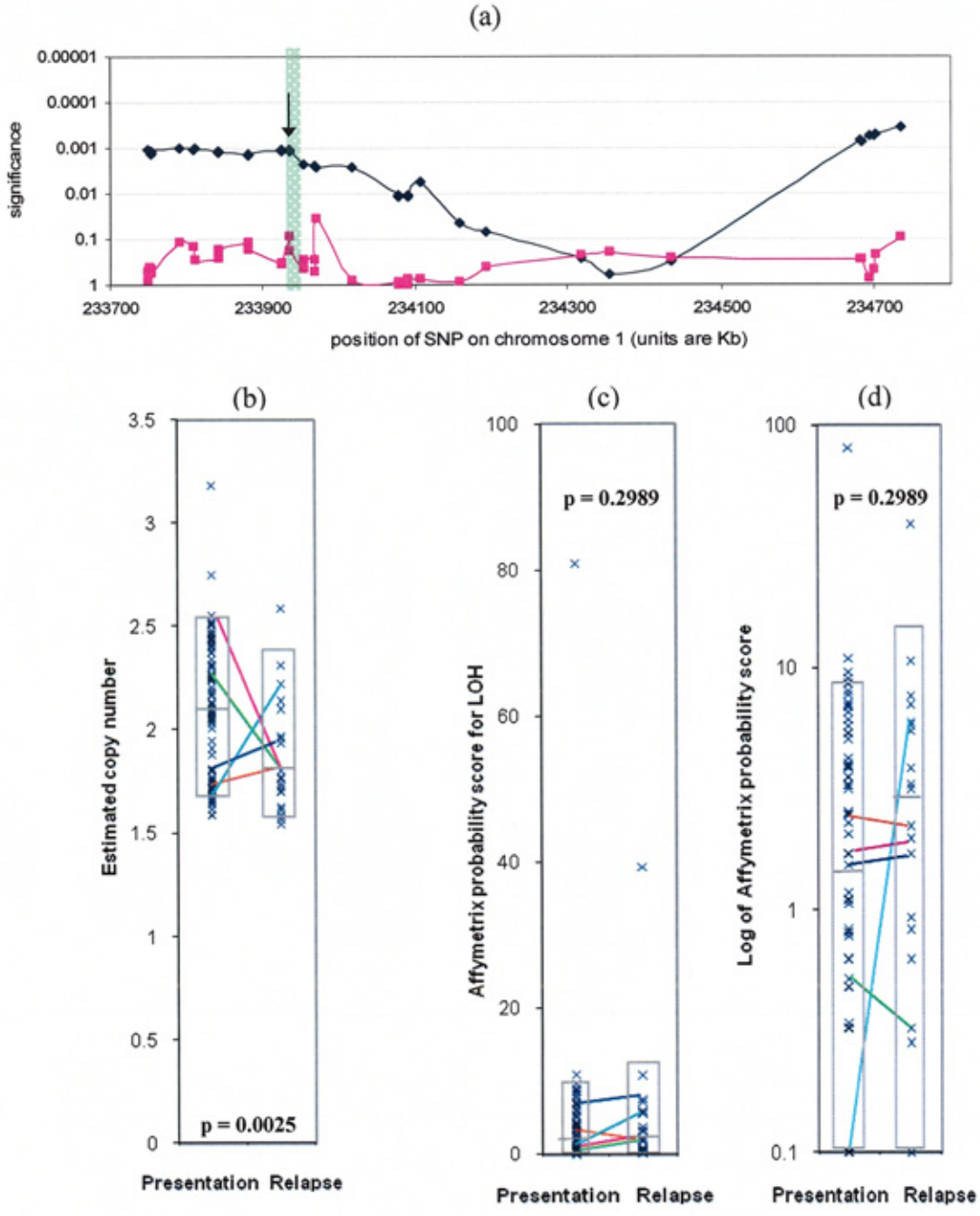
Comparing the LOH for presentation and relapse patients at this SNP did not show any statistically significant difference to indicate an overall LOH in either group (figure 4.8). Two of the presentation samples provided a probability score for LOH greater then 10 (figure 4.8c), that with the highest probability (sample L467 with a score of 26.9) was near haploid (25XY +21) in 9 metaphase spreads and is likely to be a physical loss of one copy of chromosome 1. The other, with a probability of 14.0, has normal cytogenetics and therefore suggests that LOH is likely due to other epigenetic factors (see section 4.7.14). Loss of a functioning allele could cause the majority of folates in these patients to be channelled into thymidine and purine biosynthesis and offer a degree of MTX resistance.

4.6.2 METHIONINE SYNTHASE (MTR or MS)

The region of chromosome 1 containing the methionine synthase gene shows a difference in copy number which is significant at the 0.001 level and extends 200 Kb upstream of the MS gene (figure 4.9a). This region contains genes for laminin receptor 1 (HEATR1) and actinin α_2 (ACTN2), neither of which are likely to be linked to folate metabolism or ALL and therefore are not thought to contribute to relapse.

Methionine synthase is also known as 5-methyltetrahydrofolate-homocysteine S-methyltransferase (EC 2.1.1.13) and catalyses the remethylation of homocysteine to methionine. Failure of this reaction is linked to a whole range of morbidity caused by the resulting hyperhomocysteinaemia (Goyette *et al.,* 1998) or lack of methionine (Friso *et al.,* 2002). The enzyme requires vitamin B_{12} (cobalamin) as coenzyme, so deficiency of this micronutrient will result in a drop in enzyme activity and hyperhomocysteinaemia. Loss of MS activity due to B_{12} deficiency, or a defect in the enzyme itself, will have an effect on folate metabolism because the relatively small amount of intracellar folate is no longer recycled, but gets stuck as 5-methyltetrahydrofolate. This is often referred to as the 'methyl trap' (Nijhout *et al.,* 2004) and yet, as the folate pathway '*backs up*', it is proposed that what THF is available as $N^{5,10}$methylenetetrahydrofolate is pushed toward purine and pyrimidine biosynthetic pathways and confers a degree of protection against ALL (Skibola *et al.,* 2002). Across the 33 exons on chromosome 1 that code for this enzyme (Anon, 2006c), thirteen novel mutations have been reported (Watkins *et al.,* 2002). These include deletions and nonsense mutations that generate a defective enzyme (often truncated), but the clinical consequences are related to the resulting hyperhomocysteinaemia (Kishi *et al.,* 2003; Toole *et al.,* 2004).

Figure 4.9 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 1 including the MS gene.
(b)(c) and (d) Scatterplots using data from SNP A-1643875



The scatterplot (figure 4.9b) of presentation against relapse samples for estimated copy number and probability of LOH at this position show that the gene in relapse samples has a slightly lower copy number average (ie., values closer to 2). The significant statistical difference between the presentation and relapse samples is likely

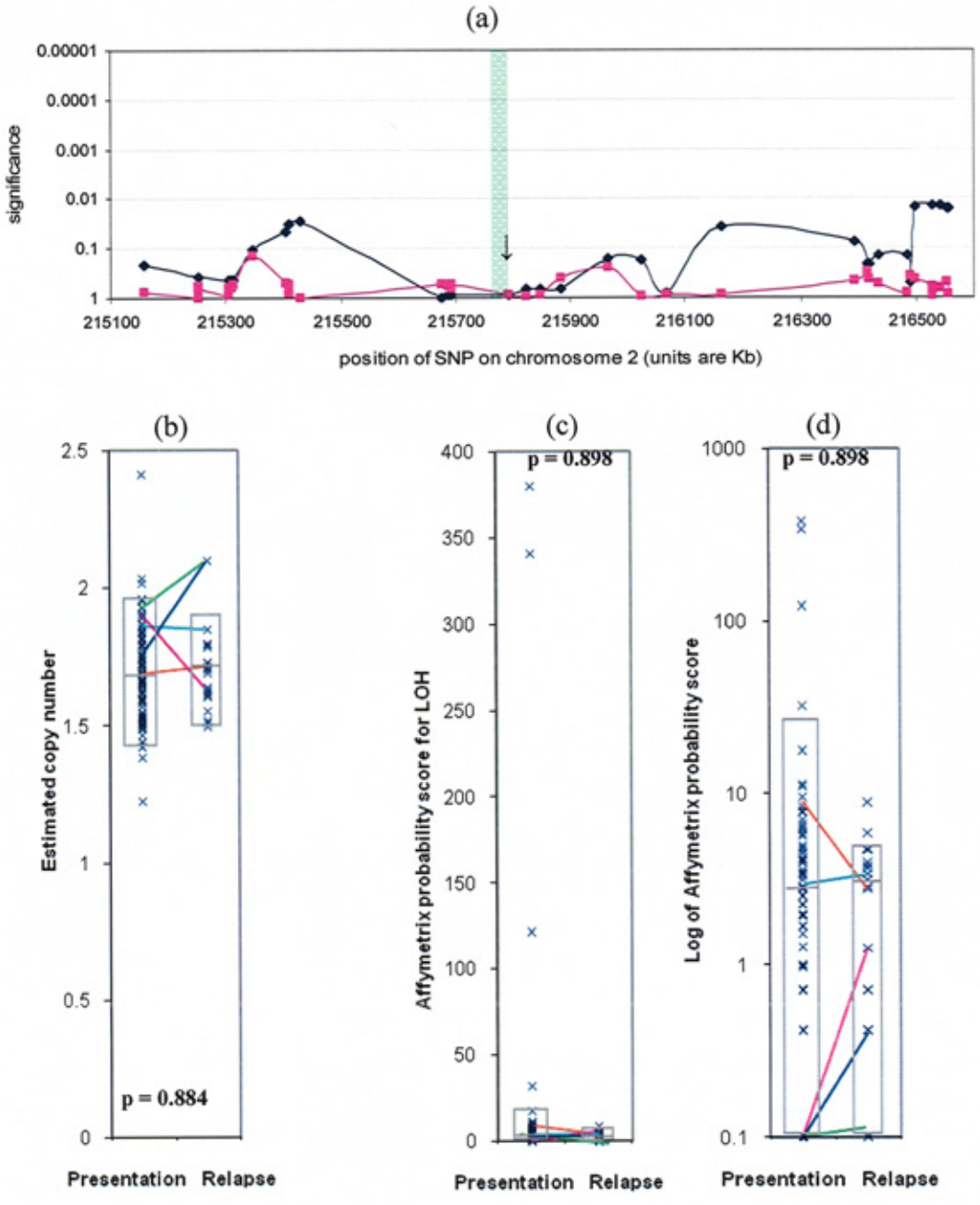
to be the result of the relatively tight clustering of both populations and does not indicate a global change in copy number across the sample groups. Looking at individual samples the presentation sample with an estimated copy number of just over three does have an extra-copy of chromosome 1. The extra copy of this gene, if expressed, would allow rapid regeneration of methionine from homocysteine, releasing folate back into the pool for purine and pyrimidine metabolism. This is only one patient and there is nothing to challenge the conclusion of Skibola *et al.,* (2002) that changes in the MS gene are unlikely to confer any degree of resistance or sensitivity to MTX.

Figures 4.9b and 4.9c show that there is one presentation sample (L467) with a very high probability for LOH, confirmed as the patient with near haploid cytogenetics and therefore a true finding. The two samples with a probability just over 10, one presentation and one relapse, have no cytogenetic findings to suggest gross changes and therefore are likely to be the result of epigenetic changes.

4.6.3 AMINOIMIDAZOLE CARBOXAMIDE RIBONUCLEOTIDE TRANSFORMYLASE (AICART)

From figure 4.10a there does not appear to be any difference between the presentation and relapse groups along the region of chromosome 2 that contains the gene for 5-aminoimidazole-4-carboxamide ribonucleotide transformylase (AICART; EC 2.1.2.3). This enzyme catalyses the penultimate reaction in *de novo* purine biosynthesis and is actually only one catalytic site on a larger multienzyme complex referred to as 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase (ATIC) (Rayl *et al.,* 1996). Plotting the data from the SNP closest to the AICART gene shows some individual sample differences.

Figure 4.10 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 2 including the AICART gene. (b)(c) and (d) Scatterplots using data from SNP A-1701154



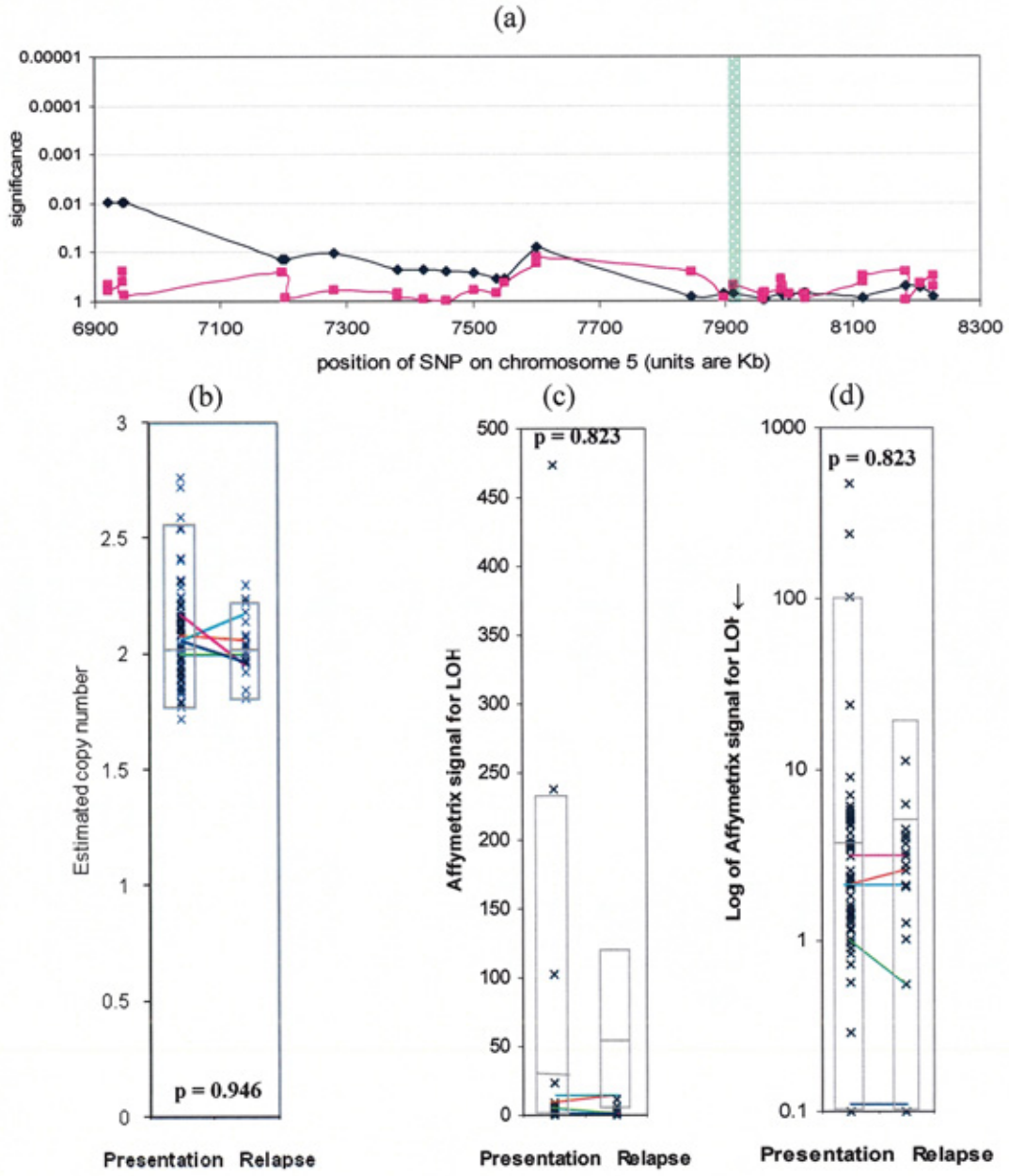
In figure 4.10b there is a presentation sample with an estimated copy number of approximately 2.5. Whilst it is impossible to have a half copy number the point is sufficiently far enough away from the tight cluster between 1.5 and 2.0 that it is would have been worthy of examining the cytogenetics for this patient, but unfortunately banding failed and there was nothing of note involving chromosome 2.

Figures 4.10c and 4.10d show three samples with probablity scores of greater than 100 for LOH. One of these, with a probability score of 341, was patient L467 previously described with near haploid cytogenetics (25XY +21). Patient L541 had a probability score of 380 but no cytogenetic results to indicate a physical loss of the whole, or part, of chromosome 2. The fact that estimated copy number for this sample was not significantly lower than other presentation samples leads to the conclusion that LOH in this case is most likely the result of epigenetic factors such as uniparental disomy suppressing one of the alleles, or a loss of such a small portion of the chromosome that this is not detectable by standard cytogenetics. The same interpretation can be made for patient L526 with a score of 121 and no apparent loss of chromosome 2.

4.6.4 METHIONINE SYNTHASE REDUCTASE (MTRR)

This enzyme is responsible for maintaining methionine synthase in a reduced state and therefore essential for catalysing the reaction described in section 4.7.2. Defects in methionine synthase reductase (MTRR; EC 2.1.1.135) have been reported, but present clinically as methionine synthase deficiency (Leclerc *et al.*, 1998). Figure 4.11a does not show an overall significant difference between the presentation and relapse groups that could account for MTX resistance and subsequent relapse. On figure 4.11b there are three samples that lie above the 95% percentile, two of these, patients L529 and L414 have extra copies of chromosome 5 confirmed by cytogenetics. However there is no extra copy evident from the cytogenetics report for sample L526, a sample with an estimated copy number comparable with these other two samples. More sensitive cytogenetic studies may be able to identify extra intrachromosomal material for this gene.

Figure 4.11 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 5 including the MTRR gene. (b)(c) and (d) Scatterplots using data from SNP A-1691055



Patients L467 and L599 had very high probability scores of 237 and 474 respectively, indicating that these two samples had a LOH. Sample L467 is the case previously mentioned with 25XY +21 and can be taken to be an appropriate signal. Sample L599 with the higher probability score did not have cytogenetic results to aid with interpretation, but is strongly suggestive of a loss of chromosomal material. The

gene lies at location 5p15.3 and the probability score is repeated for the other SNPs across the 1.3 Mb region of the p-arm. This LOH also affects the q-arm (see section 4.7.5) and suggests that one copy of chromosome 5 has been lost.

4.6.5 BETAINE METHYLTRANSFERASE (BMT)

Statistically there is little difference between the presentation and relapse samples when looking at copy number or LOH as a whole group (figure 4.12a). The estimation of copy number at SNP A-1749904 presented in figure 4.12b shows the majority of samples clustered between 2 – 2.5 and a small number of samples forming a small sub-group closer to 3.0. This would suggest that these samples have increased copy number, and for three of these four samples the cytogenetic records show an extra copy of chromosome 5 and support these findings. In figures 4.12c and 4.12d there are two presentation samples with high probability scores for LOH. L467 is the sample with near haploid cytogenetics, sample L599 unfortunately did not have cytogenetics reported and therefore could have absolute loss of chromosomal material or epigenetic suppression. However, the most striking finding is what appears to be an acquired LOH in a third patient (patient L173). The cytogenetics reported on presentation and relapse samples is the same, suggesting that LOH is probably due to an epigenetic factor. To understand this fully further work is required to confirm that the cytogenetics have not changed in the relapse sample.

To protect itself from toxic hyperhomocysteine the cell can use the transulphation pathway to metabolise and excrete the excess. However, in doing this, folate can effectively become trapped as methyltetrahydrofolate if there is no homocysteine to methylate. To increase the chance of remethylation and maintaining a balance, humans are able to transport homocysteine to the liver where the enzyme betaine

methyltransferase (BMT); also known as betaine-homocysteine S-methyltransferase (BHMT; EC 2.1.1.5) is able to recycle it back to methionine (Steenge *et al.,* 2003). This zinc-dependent enzyme is coded by the BMT gene on chromosome 5, but because it has a higher K_M for homocysteine than MTR then it only acts as a back-up mechanism when homocysteine levels are particularly high (Steenge *et al.,* 2003). The reaction is dependent on betaine as the methyl donor and provided the diet is normal, the supply of betaine and/or its precursor, choline, is rarely considered. Khohlmeier *et al* (2005) however raise the problem of low choline and betaine intake influencing folate metabolism and therefore must be considered as a potential influence on MTX resistance. It is only in dietary deficiency of betaine and choline that the SNPs found in BMT (Heil *et al.,* 2000; Meyer *et al.,* 2004) have an effect on homocysteine levels, but the full effect on folate metabolism has not been fully investigated.

Findings of this study, indicating a LOH, could cause a loss of BMT activity and make the patient more susceptible to hyperhomocysteinaemia. The associated morbidity may contribute to a poor outcome (Kishi *et al.*, 2003), but not directly confer a resistance. The opposite is true for the patients with increased copy number who may be able to reduce hyperhomocysteinaemia as a side effect and tolerate higher doses of MTX.

Figure 4.12 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 5 including the BMT (left) and the DHFR gene (right).

(a)

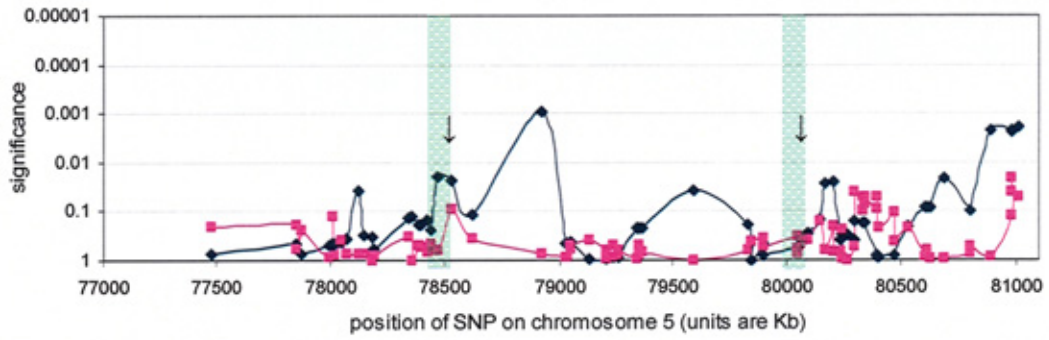
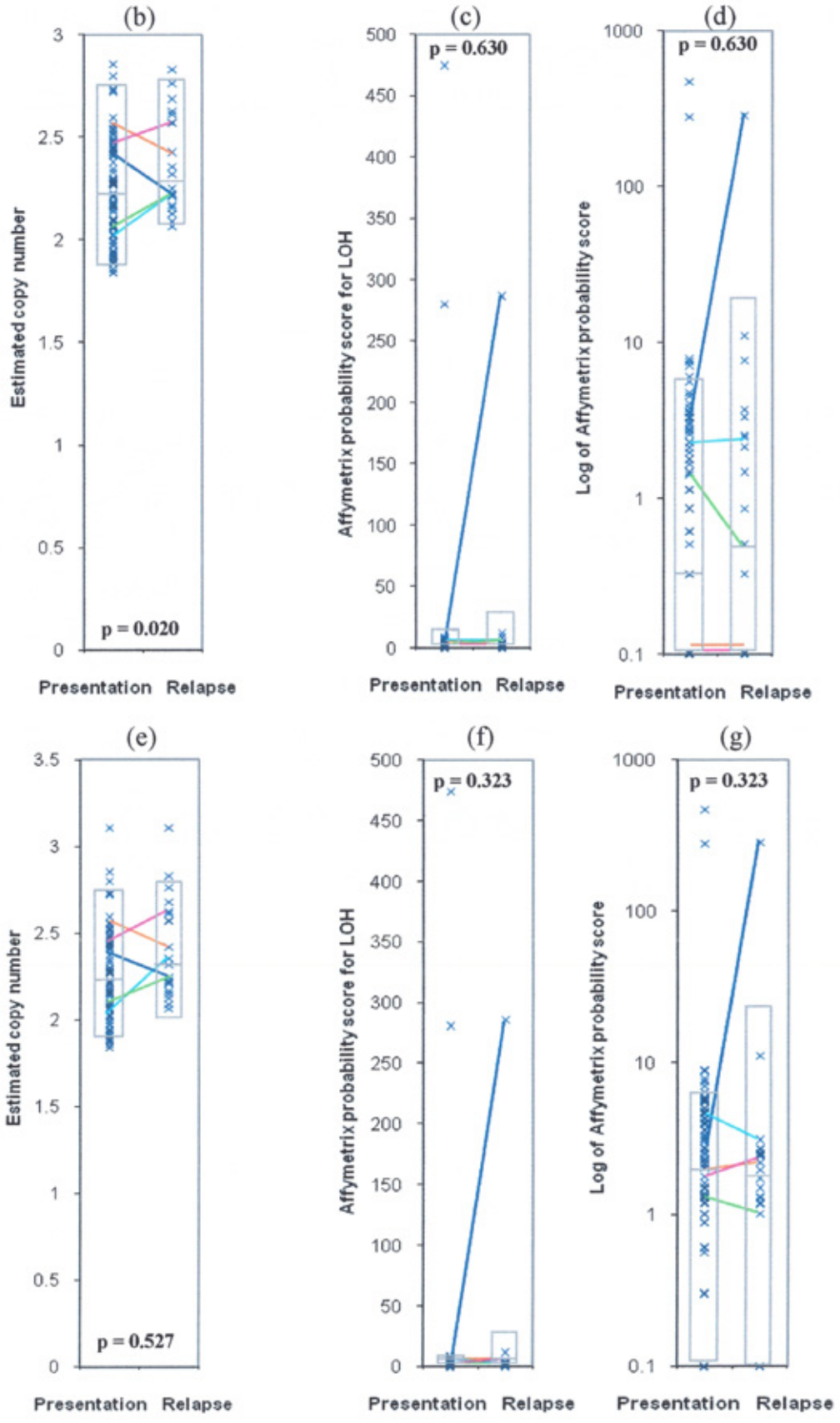


Figure 4.12 continued. (b)(c) and (d) Scatterplots using data from SNP A-1749904 for BMT; (e)(f) and (g) Scatterplots using data from SNP A-1673169 for DHFR



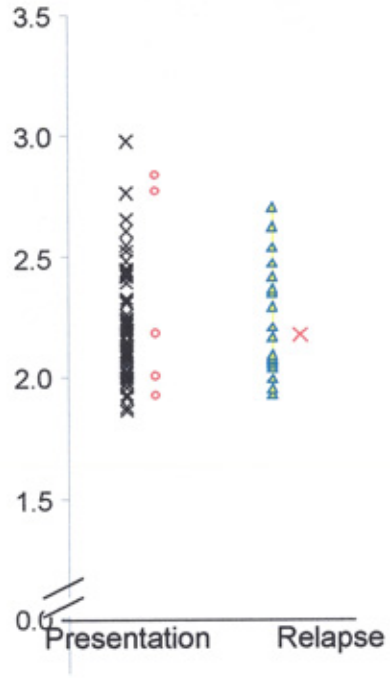
4.6.6 DIHYDROFOLATE REDUCTASE (DHFR)

It has been known for a couple of decades that increased expression of DHFR confers MTX resistance *in vitro* (Srimatkandada *et al.*, 1983; 1989; Schimke, 1988; Stark *et al.*, 1989; Goker *et al.*, 1995; Gorlick *et al*; 1996; Matherly *et al.*, 1995;1997; Iqbal *et al.*, 2000) and *in vivo* (Rots *et al.*, 2000a; Matheson *et al.*, 2007) for ALL and other cancers (Cowan *et al.*, 1982; Sowers *et al.*, 2003; Serra *et al.*, 2004;). An increase in copy number for DHFR could provide the cell with a significant advantage in developing resistance to MTX and yet none of the relapse samples had a sufficiently high probability value to indicate such a mechanism had played a part. Any increased DHFR expression is therefore likely to be due to transcription and translation of the normal copy number. It was disappointing that samples scored by cytogenetics to have an additional copy of chromosome 5 did not form the distinct sub-group with probability scores closer to 3. As can be seen in figure 4.13 the samples with confirmed +5 are distributed across the range and this reduces the confidence with which interpretation of the data can be made. It could be that having an extra copy of the gene has little impact on the expression of the enzyme because of the tight feedback control of expression by its own mRNA (discussed in section 1.10.3).

For probability of LOH shown in figures 4.12f and 4.12g, the distribution of values is almost identical to that for the neighbouring gene, BMT. With one of the relapse samples acquiring LOH and a presentation sample (L599) with a high probability for LOH, then it is possible that this could influence response to MTX. Another presentation sample with a high probability for LOH is confirmed by cytogenetics to be missing a copy of chromosome 1. Logically it could be expected that with only one allele expressing DHFR, this could result in low levels of enzyme production which in turn could easily be inhibited by relatively low levels of MTX and contribute

to a good outcome. However, it is possible that one allele would still be sufficient to respond to negative feedback control and efficiently transcribe sufficient mRNA and thus not affect the cell's sensitivity to MTX. Literature published to date has focused on transcription and translation of DHFR, as the most likely mechanism for MTX resistance in ALL and other cancer (Wade *et al.*, 1995; Guo *et al.*, 1997; Goto *et al.*, 2001; Sowers *et al.*, 2001; Dulucq *et al.*, 2008), and is unlikely to be related to gene copy number. LOH along this region of chromosome 1 could also affect the CK2α and SSBP2 genes, both reported to contribute to haematopoetic cell growth and differentiation (Wang *et al.*, 2006;Liang *et al.*, 2005).

Figure 4.13 Scatterplot showing presentation and relapse samples with (red) and without an extra copy of chromosome 5 based on cytogenetic findings.



4.6.7 FOLYLPOLYGLUTAMATE HYDROLASE (FPGH)

One of the lysosomal enzymes found in all cells, folylpolyglutamate hydrolase (EC 3.4.19.9; also known as γ-glutamyl hydrolase; GGH or γGH) catalyses the hydrolysis of glutamate residues from polyglutamylated folates and antifolate, thus facilitating

efflux from the cell via the MRP mechanism. Increasing the activity of this enzyme will reduce the time antifolates spend in the cell and is likely to confer a degree of resistance to antifolates that require polyglutamylation for maximum effect. Figure 4.14a shows no significant difference between the presentation and relapse groups as a whole.

Figure 4.14 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 8 including the FPGH gene.
(b)(c) and (d) Scatterplots using data from SNP A-1741370

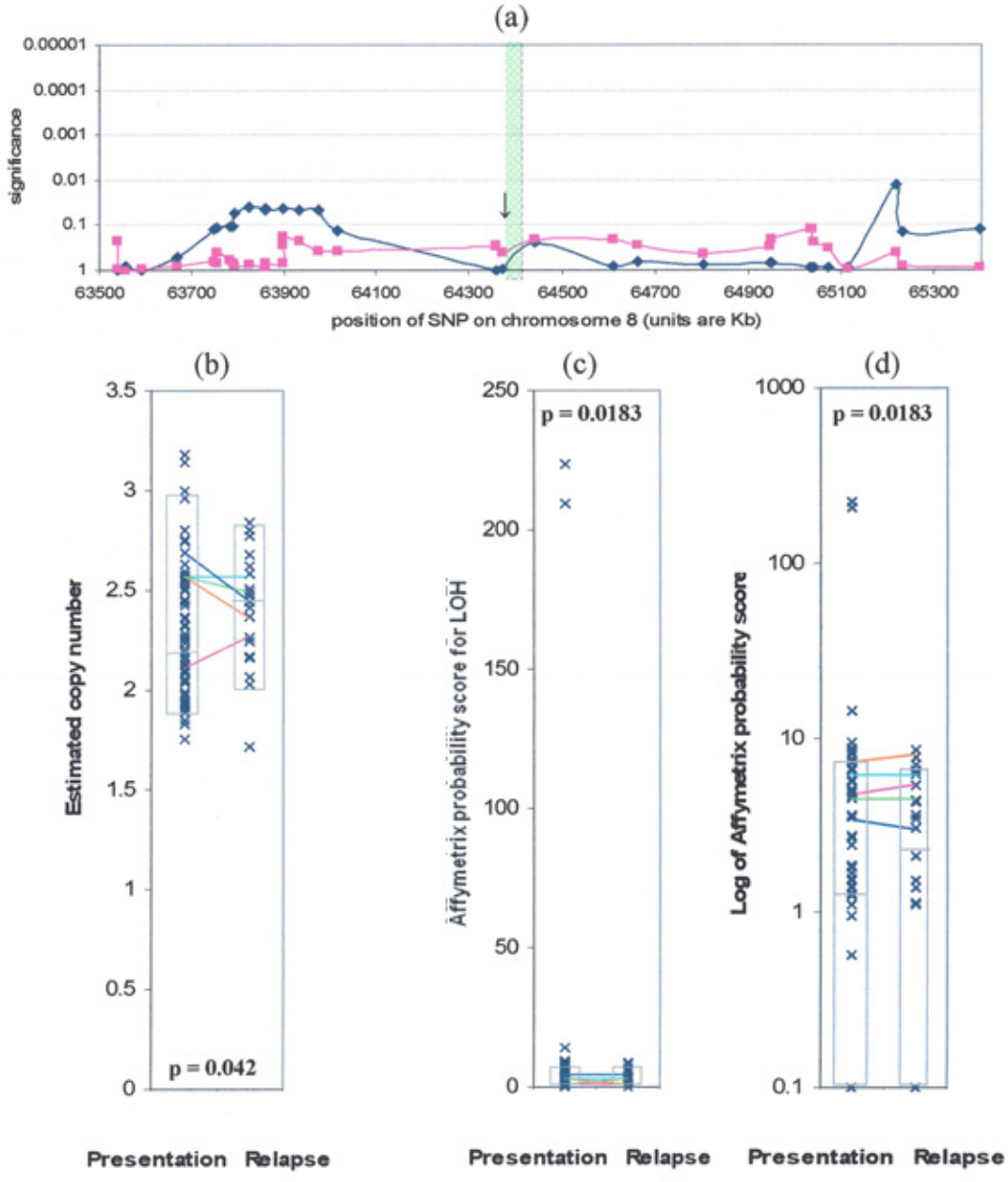


Figure 4.14b has a pattern that is now familiar for the presentation samples, ie., the majority have a normal estimated copy number and a small number of individuals have values around three and suggest an extra copy. Only four presentation samples have estimated copy number indicative of an extra copy of chromosome 8 at this location. Two of these have +8 confirmed by cytogenetics, a third has unspecified hyperdiploidy and the fourth could have intrachromosomal replication not identified by routine cytogenetics. It is disappointing that five other presentation samples with additional copies of chromosome 8 by cytogenetics have estimates for copy number of around two. Whilst this reduces the confidence in microarray results, it is still possible that these additional copies are damaged and missing the SNP location used for identification at this point on the 'chip'. If FPGH played a significant role in MTX resistance then more relapse samples could be expected with a higher copy number to facilitate excretion of unpolyglutamylated drug.

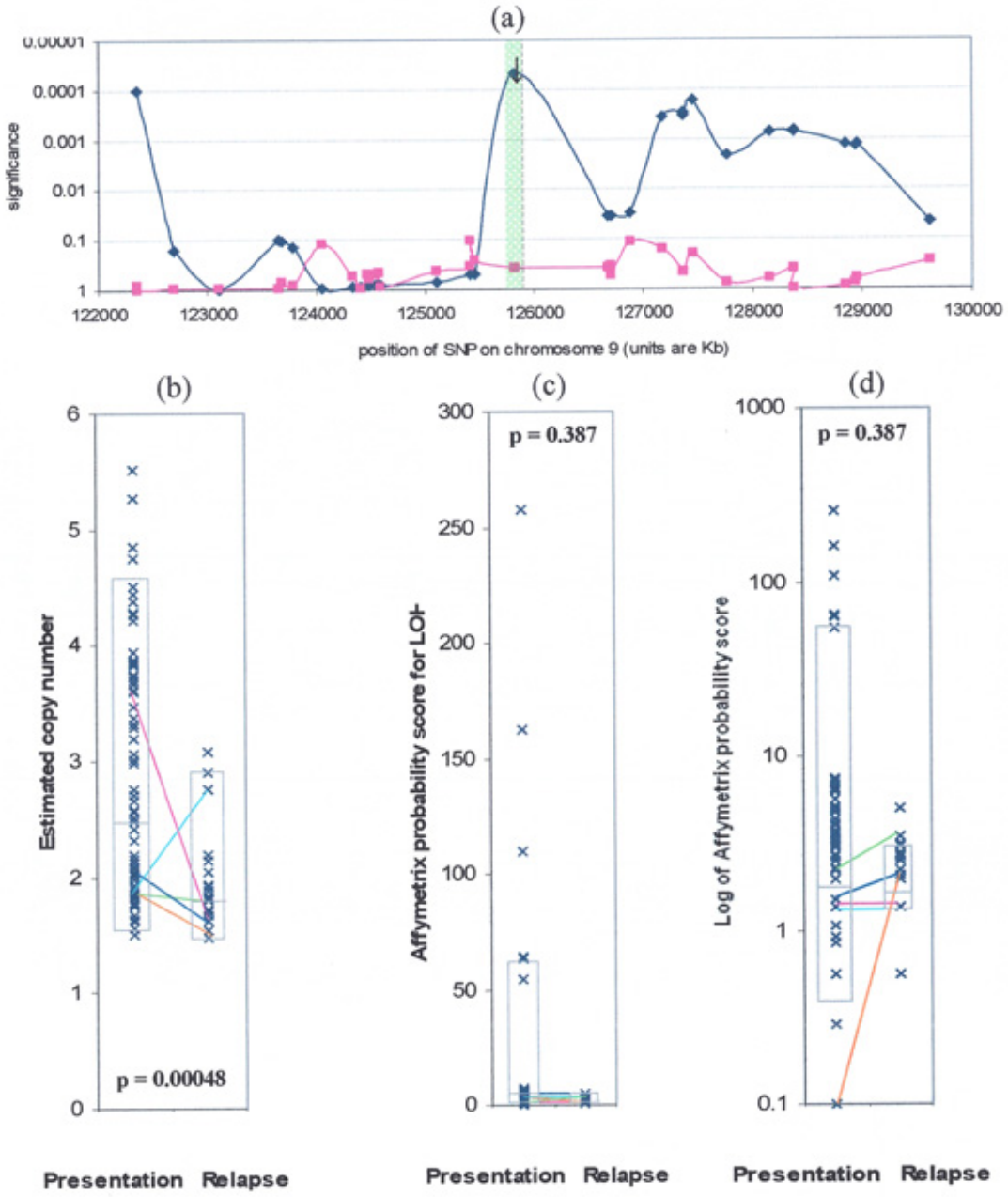
For two of the presentation samples LOH is probable. For one, there is an absolute loss of chromosome 8 in the near haploid patient and validates the probability score. The second sample did not have successful cytogenetics to confirm an absolute loss of chromosomal material or not. in the cell is required. Zheng *et al* (2001) proposed that increasing hydrolase activity is not a likely mechanism of MTX resistance, but more recently Cheng *et al* (2006) did show a relationship between FPGH activity and copy number of chromosome 8. They also showed that methylation of the promoter region of this gene plays an important role, as does the C452T SNP causing a threonine to isoleucine change in the coding sequence at residue 127 that affects binding affinity (Cheng *et al*, 2004). Therefore, for those samples with likely LOH then further work using chromatography to separate and quantitate the various intracellular polyglutamylated forms would be useful to fully understand what is happening.

4.6.8 FOLYLPOLYGLUTAMYL SYNTHETASE (FPGS)

The SNP position located on the 9q34.1 region of chromosome 9 closest to the folylpolyglutamyl synthetase gene shows a single spike in the statistical difference between presentation and relapse samples which is significant at the 0.0001 level for copy number (see figure 4.15a). The SNP distribution at this location is relatively sparse and relies on the copy number of SNP A-1703348. Analysis of the data at this point using a scatterplot (figure 4.15b) shows the expected grouping of samples for both presentation and relapse at around two, but the difference is due to a large number of the presentation samples with values greater than 3. This suggests a large number of samples with increased copy number at this location. It is unlikely to be due to increased chromosomal number as this would have shown a similar statistical difference across this whole region, but from figure 4.15a, this is clearly not the case. Three of the samples with estimated copy number greater than four have cytogenetic results confirming extra chromosome 9. The sample with the top value has two extra copies and this would indicate that the microarray is capable of providing results that allow copy number to be estimated effectively.

The bimodal distribution of presentation samples seen in figure 4.15c is similar to that seen for MTHFR (section 4.7.1) but in the opposite direction. Interestingly there is a relationship between the results for the two locations, ie., if the MTHFR value is low the FPGS estimated copy number is normal, if the MTHFR value is normal then that for FPGS number is high. From the data available no theory has been proposed to explain this finding.

Figure 4.15 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 9 including the FPGS gene. (b)(c) and (d) Scatterplots using data from SNP A-1703348



It is known that cells which lack FPGS activity have a degree of antifolate resistance (Zeng *et al.*, 2001; Liani *et al.*, 2003) and thus the presentation patients with a higher copy number may respond better to treatment and could account for why few are noted in the relapse group. The enhanced polyglutamylation of MTX following presentation could ensure maximum pharmacological effect of the drug. Therefore, it

could be expected that a LOH may lead to more effective resistance, and yet, the patients with a high probability of LOH are not found in the relapse group (figures 4.15c and 4.15d).

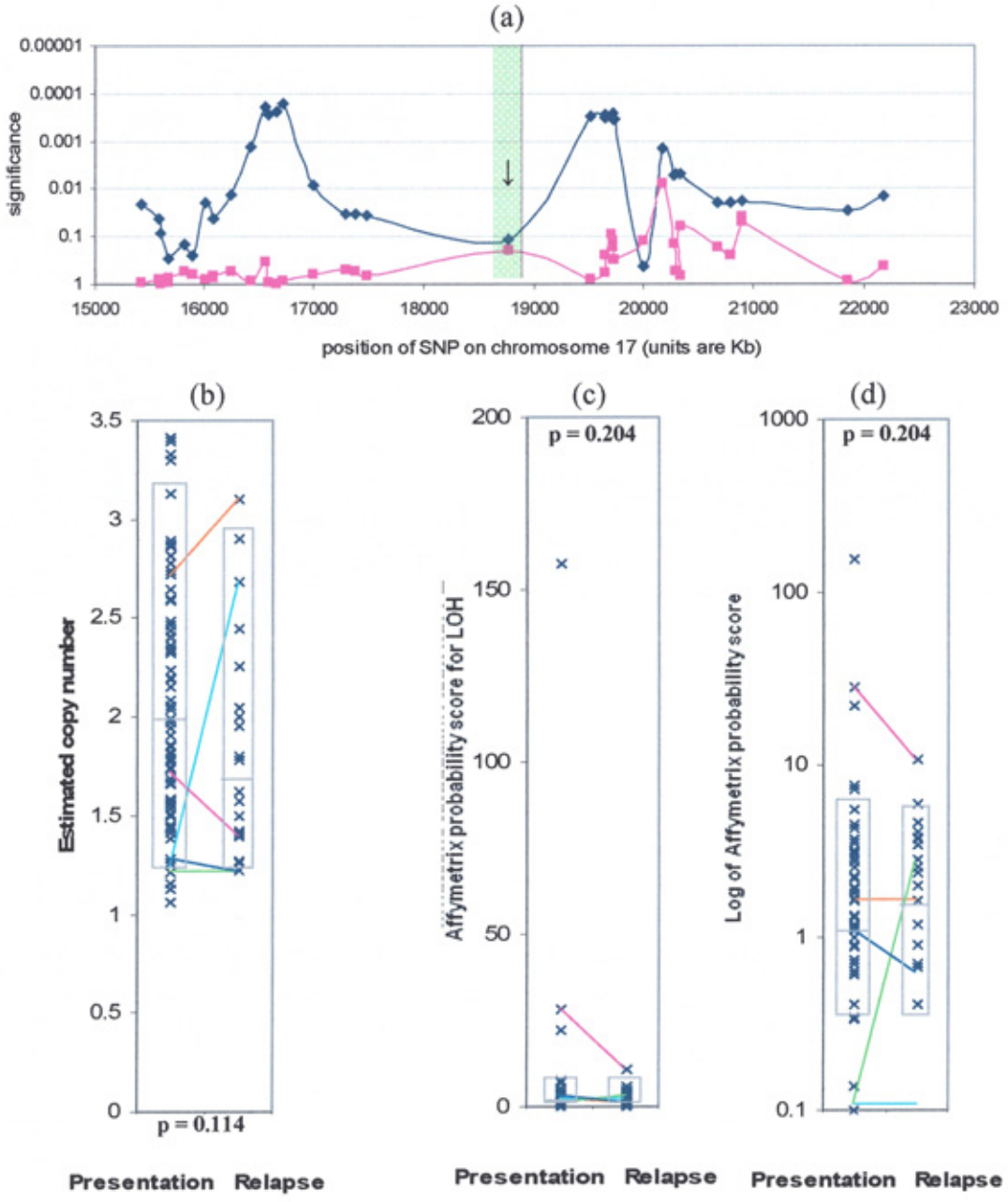
4.6.9 SERINE HYDROXYMETHYL TRANSFERASE (SHMT)

This enzyme fails to get the attention that others have (Renwick *et al.*, 1998; Rao *et al.*, 2000), and yet it is crucial for the methylene transfer to folate in the first place (figures 1.4 and 1.5). Serine hydroxymethyl transferase (SHMT; EC 2.1.2.1, also known as glycine hydroxymethyltransferase) has two isoenzymes; cytoplasmic (cSHMT) and mitochondrial (mSHMT) forms coded by SHMT1 and SHMT2 genes, on chromosomes 17 and 12 respectively. Although not the only mechanism to transfer 1C moieties to folate, it is considered to be the rate limiting step in shunting folates either into methionine regeneration or along *de novo* purine metabolism (Herbig *et al.*, 2002; Skibola *et al.*, 2004). From figure 4.16a the nearest SNP location to the cytoplasmic serine hydroxymethyl transferase gene on chromosome 17 shows insignificant difference in copy number between the presentation and relapse groups. Downstream is a region of significant difference that may influence the SHMT gene but is considered unlikely and has not been investigated any further here. The scatterplot for estimated copy number (figure 4.16b) shows a wide distribution of values either side of two and other than the small number of samples with values greater than 3 does not appear to be significant. Samples L500, L599, L600, L601 and L659 did not have cytogenetics to confirm extra copy number as is suggested by microarray estimated copy number.

The relapse sample that stands out in figure 4.16c for LOH is L467. This sample clearly has missing chromosomes (25XY+21), including 17, and it would be expected

that this sample would have a very low estimated copy number. Whilst this falls within the lower quartile, it is not significantly lower at 1.5 than the rest. The other two samples with a probability score greater than 10 could be considered significant because one of them (L529) has hypodiploid cytogenetics with a likely loss of chromosome 17.

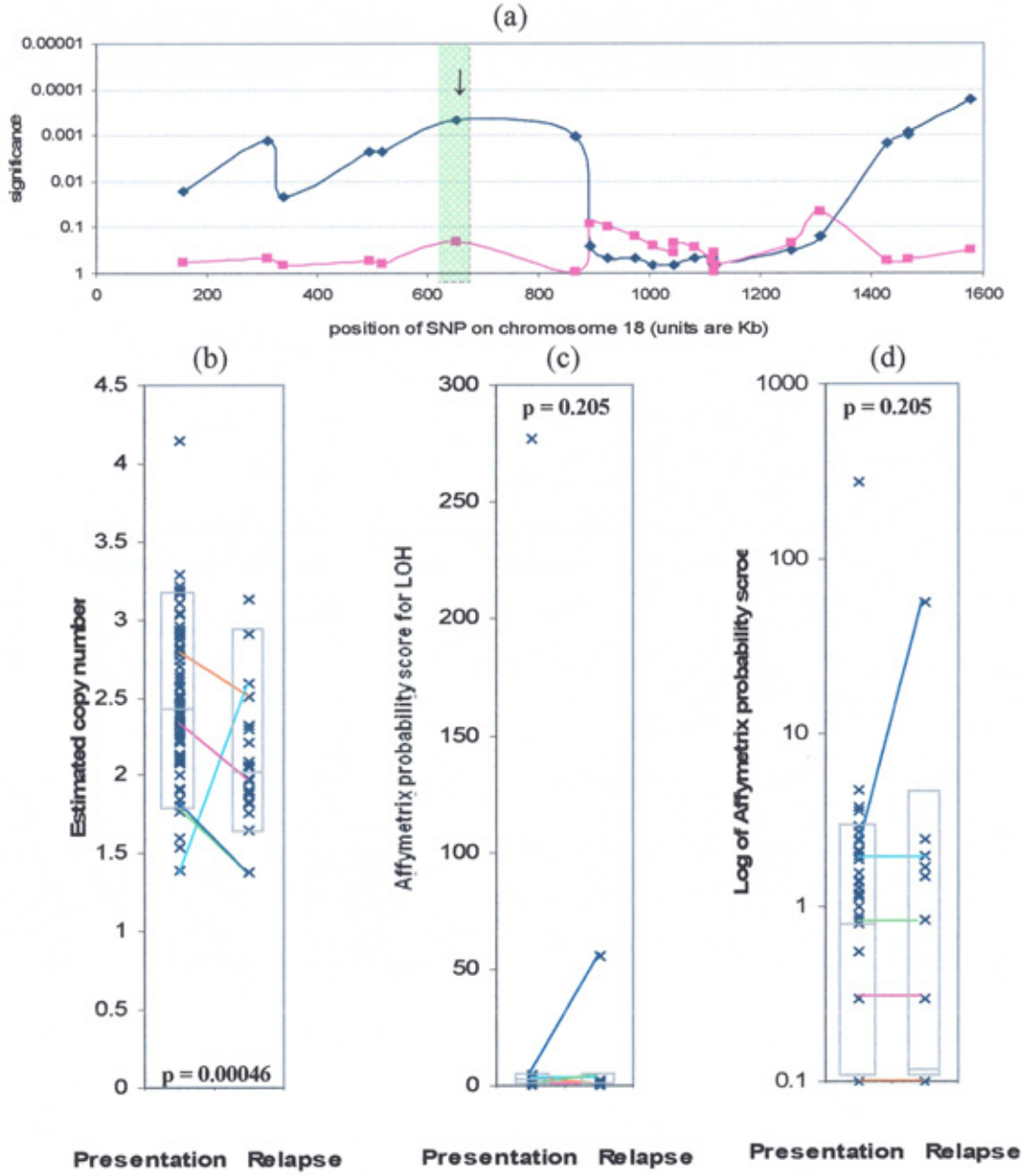
Figure 4.16 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 17 including the SHMT gene.
(b)(c) and (d) Scatterplots using data from SNP A-1659463



4.6.10 THYMIDYLATE SYNTHASE (TS)

Figure 4.17a shows that for copy number there is a region of interest at the start of the p-arm of chromosome 18 that includes the thymidylate synthase gene. The p-value of <0.0001 is highly significant and the scatterplot (figure 4.17b) shows that this is likely to be due to the slightly higher values for the presentations samples which are clustered between values of 2 to 3, but the actual difference in signal between the two groups is minimal. Sample L519 is scored +18 on cytogenetics and supports the high estimated copy number, although a value of 4.1 would suggest even more copies. Figures 4.17c and 4.17d have the near haploid sample clearly differentiated from the other presentation samples, but the most significant finding is for sample L173. This sample appears to have acquired a loss of genetic material between presentation and relapse. This is unlikely to be an epigenetic suppression of one allele because the copy number in figure 4.17b has dropped from 1.85 to 1.38 (or 2 to 1 if taken to nearest whole number). As there is no obvious cytogenetics reported to account for this then is could be the result of partial loss of the p-arm. The impact that the VNTR could have on expression of TS and interpretation of the data above needs further investigations. A simple PCR of gDNA from normal and relapse patients using primers either side of the VNTR will generate two products of different size. These can be visualised and scored using distance travelled on an agarose electrophoresis gel. DHPLC could also be employed to differentiate the 2R and 3R variants and if optimised appropriately should also be able to pick up the G/C SNP in the second VNTR element. With this data a more complete interpretation can be made about the impact on relapse and MTX resistance.

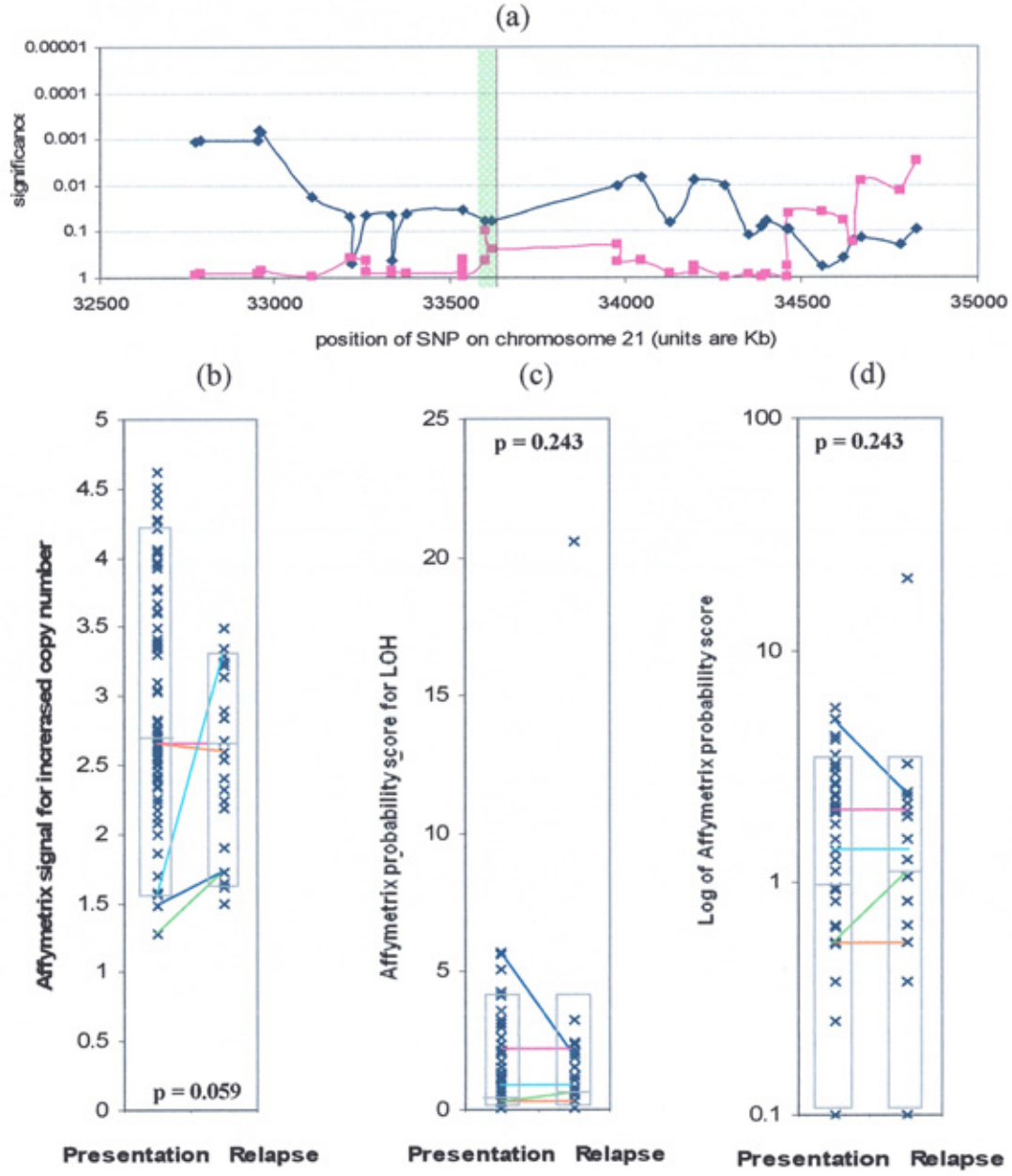
Figure 4.17 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 18 including the TS gene.
(b)(c) and (d) Scatterplots using data from SNP A-1731420



4.6.11 GLYCINAMIDE RIBONUCLEOTIDE TRANSFORMYLASE (GART)

Purine biosynthesis involves two folate dependent enzymes which are also inhibited by MTX and thus further potentiate the anticancer role of the drug. The first of these is glycinamide ribonucleotide transformylase (GART; EC 6.3.4.13) which catalyses

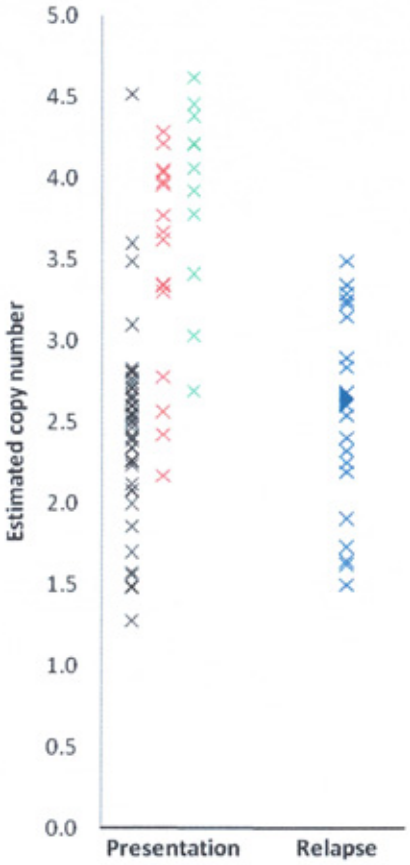
Figure 4.18 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 21 including the GART gene. (b)(c) and (d) Scatterplots using data from SNP A-1668065



the conversion of glycinamide ribotide to formylglycinamide ribotide using N^{10}-formyltetrahydrofolate as the carbon donor. The coding sequence for this enzyme is located to the GARS locus found on chromosome 21, which when fully elucidated was shown to code for a multifunction protein with three different catalytic sites (Daubner *et al.*, 1986; Schild *et al.*, 1990; Brodsky *et al.*, 1997). Each of the catalytic

sites perform a different reaction in the *de novo* synthesis of purines, but having them linked as one structure means that steps of the pathway are intrinsically balanced (1-to-1-to-1). The GART region resides toward the N-terminal end and was shown to be inhibited by MTX (Daubner *et al.*, 1986). Figure 4.18a shows that there is no overall difference between the two groups for either estimated copy number or probability of LOH. Whilst there is some overlap, figure 4.18b shows a cluster of presentation samples with estimated copy number greater then three. This reflects the high frequency of one or more extra copies of chromosome 21 in ALL. Sub-classification of the presentation samples using cytogenetics information and re-plotting the data (figure 4.19) shows that the estimated copy number reflects hyperdiplody (this is even clearer for other genes on this chromosome).

Figure 4.19 Scatterplot of presentation and relapse samples at SNP A-1668065 using cytogenetics information (from left to right, normal cytogenetics for chromosome 21 (x), one extra copy +21 (x), two or more extra copies (x), relapse samples (x).



Under normal conditions having increased copy number of this gene is unlikely to have any effect on purine synthesis because GART catalyses an intermediary reaction down-stream of the rate-limiting step. However, in the presence of MTX an increased copy number may lead to increased expression of the enzyme and thus overcome inhibition. For this to be a real advantage to the cell it must also have a simultaneous mechanism to avoid inhibition of the other more sensitive targets of MTX (ie., DHFR).

With the exception of patient L296 in the relapse group, figures 4.18b and 4.18c indicates nothing of note regarding LOH for this region of chromosome 21. No cytogenetics data was available for patient L296 to determine if there was any loss of chromosome 21. A loss of chromosome 21 is associated with poor prognosis in a number of cancers (Raimondi *et al.,* 1992; Whitehead *et al.,* 1992; Ito *et al.,* 1999; Belkov *et al.,* 1999) and in ALL, an increased chance of relapse (Chessels *et al.,* 2001; Taub and Ge, 2005).

4.6.12 CYSTATHIONINE β-SYNTHASE (CBS)

The 21q22.3 region of chromosome 21 shows that copy number is significantly different between the presentation and relapse samples (see figure 4.18a). This region includes the CBS gene; 30 Kb upstream is a gene for the PKNOX1 transcription factor and 30Kb downstream is the U2AF1 gene that codes for part of the splicesome. Whilst either of these could have some influence on the cell's resistance via effects on transcription of folate associated genes, or alternative splicing of proteins such as RFC, further work would be required to confirm a link with folate metabolism.

The highly significant p-value obtained for the difference between relapse and presentation samples seen in figure 4.20a is due to the group of presentation patients who have extra copies of chromosome 21. The subgroup of patients with estimated copy number greater than three seen in figure 4.20b can be more clearly identified as those with extra copies in figure 4.21.

Figure 4.20 (a) Significance of difference between presentation and relapse samples at SNP positions across chromosome 21 including the CBS (left) and the RFC gene (right). (b)(c) and (d) Scatterplots using data from SNP A-16877757 for CBS

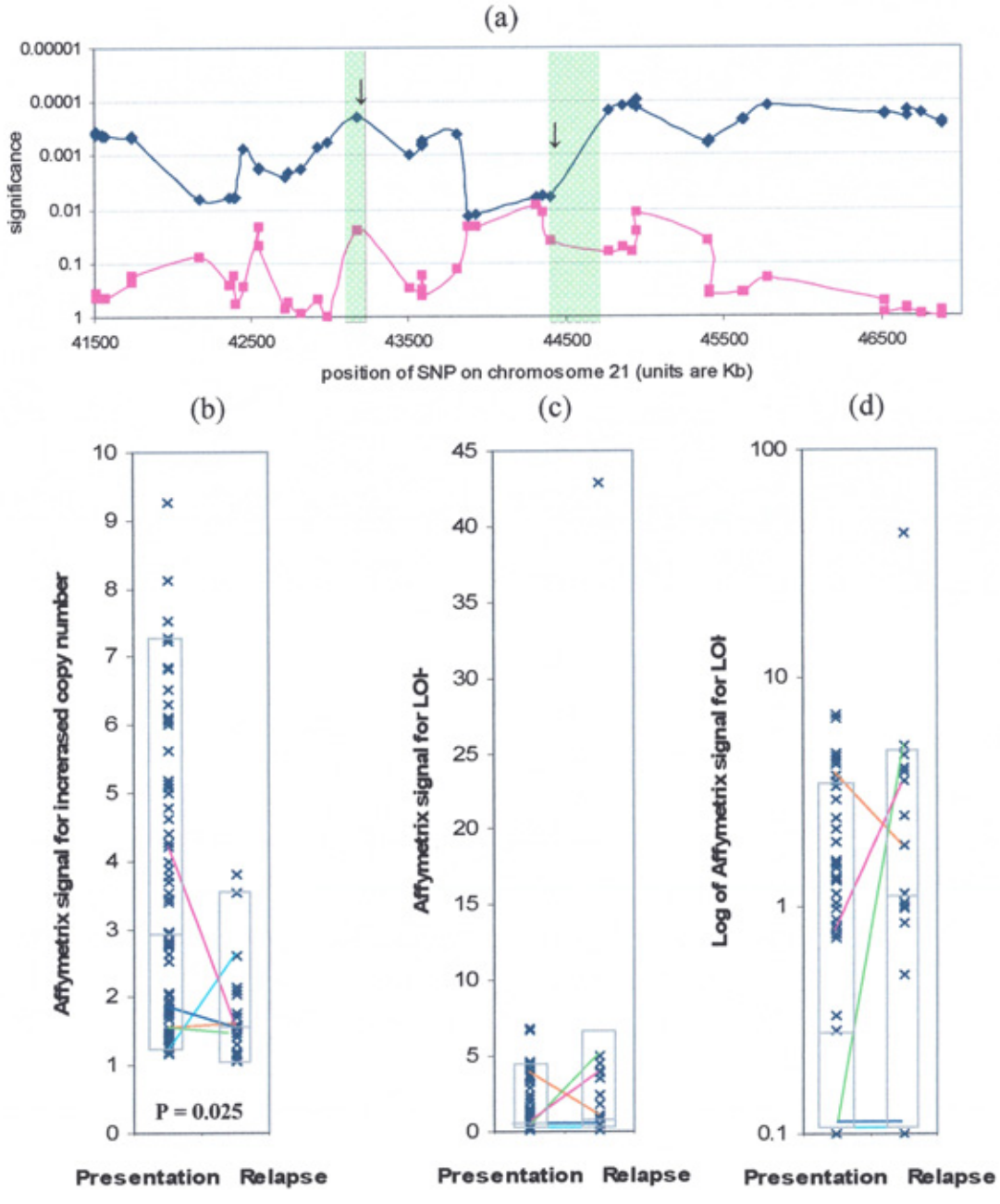
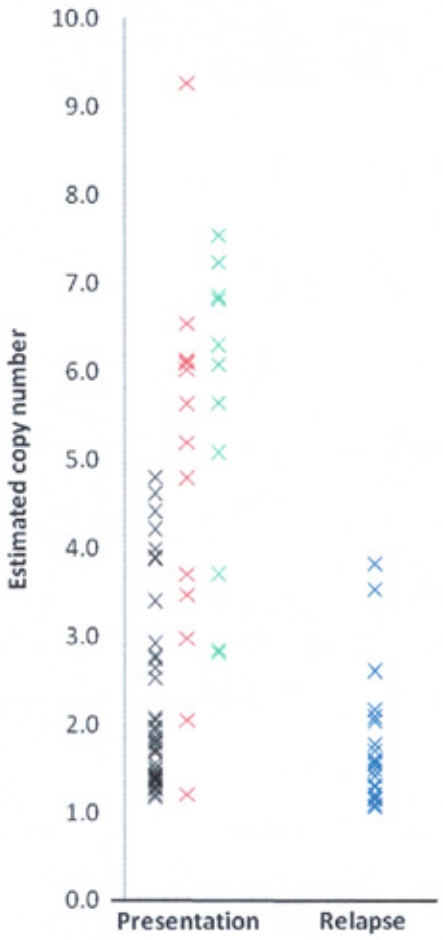


Figure 4.21 Scatterplot of presentation and relapse samples at SNP A-16877757 using cytogenetics information (from left to right, normal cytogenetics for chromosome 21 (x), one extra copy +21 (x), two or more extra copies (x), relapse samples (x).



Homocysteine sits at a branch-point in methionine and cysteine metabolism. Under normal circumstances about half of the homocysteine is remethylated back to methionine for use in protein synthesis or as a methyl donor, but the remainder is irreversibly converted to cystathionine and then cysteine by cystathionine β-synthase (CBS; EC 4.2.1.22). This second pathway is referred to as the 'transulphation pathway'. Which route homocysteine takes is influenced by the levels of *S*-adenosyl methionine; if low, remethylation is favoured to conserve methionine; if high, the cell can afford to expend methionine and transulphation is favoured.

Using information obtained from studying Down's syndrome, which has increased cellular activity of this enzyme (Fillon-Emery *et al.,* 2004; Taub and Ge, 2005), it is not surprising that intracellular levels of homocysteine and *S*-adenosyl methionine (SAM) are lower. The latter is a known allosteric inhibitor of MTHFR and thus influences whether folates are shunted to pyrimidine metabolism or allowed to continue to homocysteine remethylation (Ito *et al.,* 1999; Taub and Ge, 2005). The presentation samples with higher copy number and therefore potentially increased expression of this enzyme would follow the pattern seen in Down's syndrome, ie., loss of MTHFR inhibition would decrease the amount of folate going into synthesis of thymidine and increase the sensitivity of the cell to the action of MTX. The result is that extra copies of chromosome 21 are a good prognostic indicator.

With the exception of sample L296, there is no suggested LOH of this gene associated with relapse. L296 has a high value across the whole of the q-arm of chromosome 21 and suggests a gross loss of genetic material, but the cytogenetics for this patient failed and therefore cannot be commented on further.

4.6.13 REDUCED FOLATE CARRIER

Like CBS, the significant P-value for copy number between presentation and relapse samples is due to subpopulations forming clusters between 3–4 and then again 4-5 which do not appear in the relapse samples (figures 4.20b and 4.22a). Figure 4.23 shows even more clearly that estimated copy number derived from the microarray data correlates with actual copy number determined by cytogenetics.

Figure 4.22. (a)(b) and (c) Scatterplots using data from SNP A-1754481 for RFC.

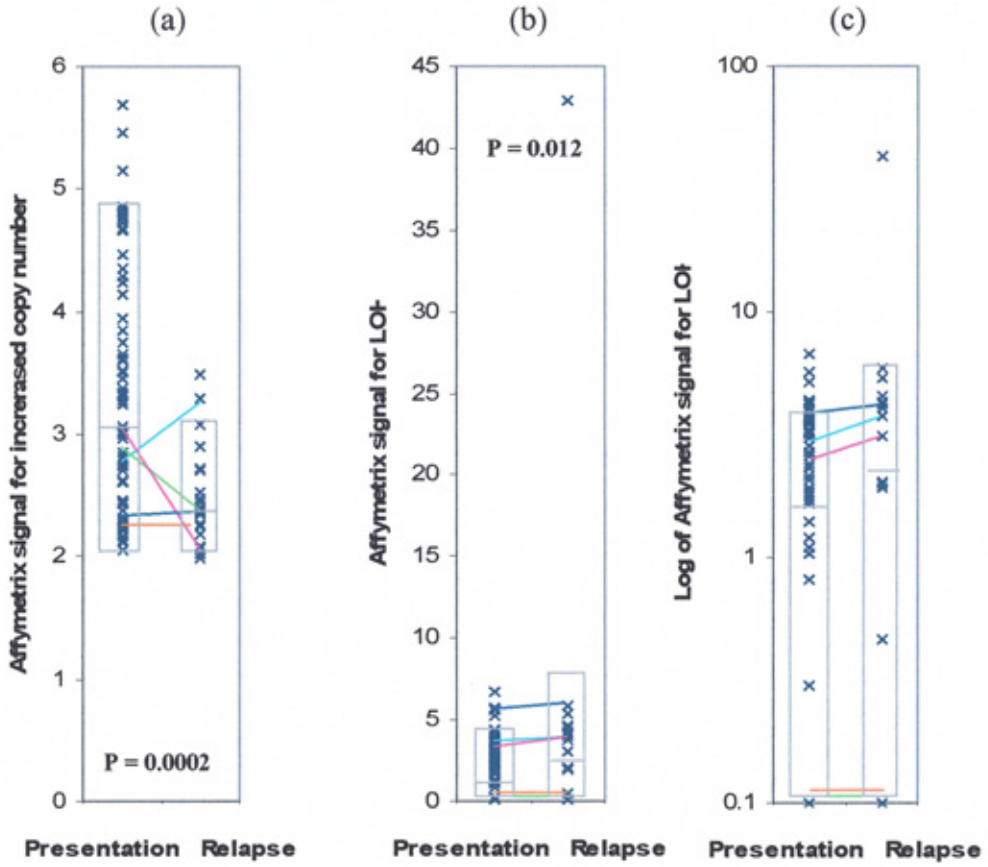
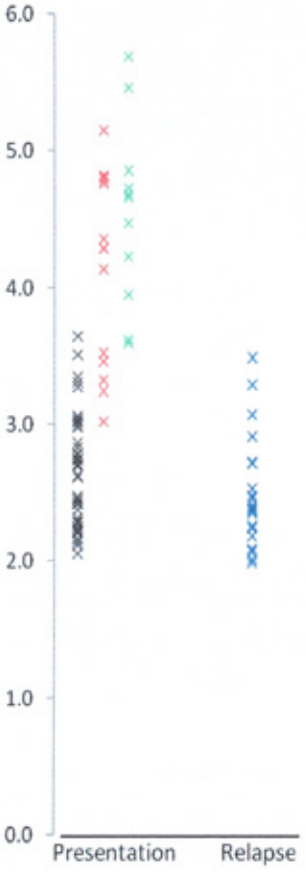


Figure 4.23 Scatterplot of presentation and relapse samples at SNP A-1754481 using cytogenetics information (from left to right, normal cytogenetics for chromosome 21 (x), one extra copy +21 (x), two or more extra copies (x), relapse samples (x).



Follow-up studies will be able to determine if the children with high copy number ultimately had better long-term survival and thus confirm the successful outcome associated with hyperdiploidy involving chromosome 21 (Taub and Ge, 2005; Moorman *et al.*, 2008).

If an increased copy number improves sensitivity to MTX, then it would be expected that a loss of chromosome 21, or at least this gene, would provide a degree of resistance (Moscow *et al.,* 1995; Gong *et al.,* 1997; Gorlick *et al.,* 1997; Zhang *et al.,* 1998c; Trippett & Bertino, 1999; Sadlish *et al.,* 2000; Ding *et al.,* 2001b; Rothem *et al.*, 2003). There was little evidence in this study for such loss, but a more subtle loss of gene function could be detected from the LOH data. Only one of the relapse patients (L296) had a significantly increased probability that this had occurred. Cytogenetics was not available on this patient to rule out any gross changes, but the copy number value of 2.0 for this patient suggests that this is unlikely. The LOH is therefore likely to be an epigenetic factor.

4.6.14 ANALYSIS OF PAIRED PRESENTATION AND RELAPSE SAMPLES

Five of the samples had paired presentation/relapse data which was examined for common trends. Estimated copy number for pairs L173, L181 and L479 tended to remain close to a value of 2.0 and any slight variation was taken to be the result of acceptable analytical error (table 4.3). This was interpreted to mean that copy number had not changed between presentation and relapse.

Patient L83 has a tendency for estimated copy number to increase, which was particularly noticeable, ie., doubled, for MTHFR (figure 4.8b), SHMT (figure 4.16b), TS (figure 4.17b), GART (figure 4.18b) and CBS (figure 4.20b). In looking for a

relationship between these five enzymes that would suggest a more effective mechanism for MTX resistance, the first three are linked by N^5,N^{10}-methylenetetrahydrofolate (figure 1.5) and GART only one reaction away. If there was increased expression of these four enzymes then THF would be loaded with 1-C and ready for use in general methylation reactions, pyrimidine and purine synthesis, in theory providing effective folate metabolism. However, this would all depend on DHFR being able to reduce DHF to THF in the first place, and as the primary inhibitory action of MTX is on DHFR, then the cell would have to overcome this problem first. There is no evidence of increased copy number for DHFR in this study, but as transcription and translation of this gene are common findings in MTX resistance (Srimatkandada *et al.*, 1983; 1989; Schimke, 1988; Stark *et al.*, 1989; Goker *et al.*, 1995; Gorlick *et al*; 1996; Matherly *et al.*, 1995;1997; Iqbal *et al.*, 2000; Rots *et al.*, 2000a; Matheson *et al.*, 2007) then it is possible that this patient's relapse is the result of these cumulative changes leading to an effective resistance.

Table 4.3 Estimated copy number for paired samples; presentation (A); relapse (C)

	Sample 83		Sample 173		Sample 181		Sample 479		**Sample 554**	
SNP ID	**A**	**C**	A	C	A	C	A	C	**A**	**C**
A-1721350	**0.27**	**1.22**	0.49	0.49	0.41	0.49	1.77	2.18	**2.93**	**1.12**
A-1643875	1.69	2.23	1.75	1.97	2.34	1.77	1.72	1.77	2.51	1.8
A-1701154	1.89	1.85	1.74	2.10	1.9	2.01	1.69	1.7	1.84	1.66
A-1691055	2.08	2.29	2.03	1.97	1.99	1.97	2.21	2.08	2.18	1.97
A-1749904	2.02	2.23	2.44	2.21	2.09	2.63	2.56	2.43	2.43	2.57
A-1673169	2.52	2.42	2.42	2.62	2.98	2.62	2.09	2.09	2.13	2.07
A-1741370	2.59	1.72	2.63	2.47	2.55	2.47	2.56	2.37	2.1	2.26
A-1703348	1.89	2.75	2.01	1.83	1.85	1.87	1.85	1.55	**3.6**	**1.71**
A-1659463	**1.27**	**2.26**	1.38	1.22	1.21	1.22	2.72	3.1	1.78	1.47
A-1731420	**1.4**	**2.59**	1.85	1.38	1.81	1.88	2.8	2.51	2.35	1.98
A-1668065	**1.58**	**3.34**	1.49	1.64	1.28	1.64	2.66	2.6	2.64	2.68
A-1687757	**1.21**	**2.6**	1.85	1.41	1.44	1.06	1.5	1.54	**4.21**	**1.48**
A-1754481	2.74	3.29	2.24	2.39	2.86	2.35	2.21	2.24	3.04	2.06

Using a probability score greater than 10 as significant for LOH, then of the five paired samples only patient L173 provided results for further discussion (table 4.4). A highly significant probability for LOH was seen for BMT (figure 4.12c), DHFR (figure 4.12f), changing from 4.01 to 286 and 2.31 to 286 respectively, and to a lesser degree TS (figure 4.17c) which changed from 8.71 to 55.7. The SNPs closest to BMT and DHFR both reside on chromosome 5 and therefore would suggest a loss of this chromosome (to be confirmed by cytogenetics on the relapse sample). Unfortuantely, there was no additional cytogenetics information on the relapse sample to help determine if LOH at the location closest to TS is gross loss of material or the result of epigenetic factors. It is difficult to understand how LOH for these three enzymes will contribute to MTX resistance in these relapsed patients.

Table 4.4 Probability of LOH for paired samples; presentation (A); relapse (C).

	Sample 83		**Sample 173**		**Sample 181**		**Sample 479**		**Sample 554**	
SNP ID	A	C	A	C	A	C	A	C	A	C
A-1721350	0.54	0	0	0	0.62	0.33	1.89	1.89	3.02	3.02
A-1643875	0	7.69	1.71	1.96	0.52	0	3.27	3.33	1.71	1.71
A-1701154	2.88	2.88	0	0	0.98	0.71	2.52	2.79	4.71	4.71
A-1691055	3.43	3.72	0	0.56	0	0	9.11	2.57	0	1.03
A-1749904	3.09	3.34	**4.01**	**286**	1.14	0.87	0	0	0	0
A-1673169	5.61	2.66	**2.31**	**286**	1.32	1.02	2.45	2.72	2.16	3.08
A-1741370	6.48	6.48	4.83	4.25	5.05	5.29	7.21	7.81	5.71	6.18
A-1703348	1.95	2.24	2.23	3.1	3.14	5.01	0	2.24	2.24	2.24
A-1659463	0.04	0.04	1.16	0.66	0	4.11	1.2	1.2	22.3	10.6
A-1731420	2.48	2.48	**8.71**	**55.7**	0.85	0.85	0	0	0.3	0.3
A-1668065	1.25	1.25	5.7	2.46	0.65	1.07	0.66	0.66	2.19	2.19
A-1687757	0	0	0	0	0	0	4.2	1.83	0.83	4.04
A-1754481	3.37	4.28	5.71	5.9	0	5.9	0	0	3.21	4.04

4.6.15 HIGH PROBABILITY FOR LOH

Without specifically targeted techniques such as FISH, cytogenetics scoring of metaphase spreads is a relatively coarse technique for picking up small changes in the DNA. Microarray-comparative genomic hybridization (CGH) allows screening of the genome with increasing sensitivity as the 'chip' density increases (Anon, 2005; Meaburn *et al.,* 2006; Ryder S, 2007; Kutchinskaya *et al.,* 2008). As smaller and smaller regions of the genome can be screened, then it is possible that LOH, as suggested in a number of samples in this study, is due to small alterations at the SNP locations. Comparative levels of sensitivity using traditional techniques would require numerous location specific, time consuming, and therefore expensive investigations to be performed. But now, further investigations to determine if there are structural losses or gains to the chromosome can be more efficiently conducted.

Samples with high probability for LOH and normal estimated copy number suggests epigenetic factors are involved, such as methylation and uniparental disomy (UPD). UPD is now an accepted mechanism for preferential expression of one allele in the cell. In the majority of genes this does not cause a problem because the favoured allele codes for a fully functional protein. But if the preferred allele is defective then this may cause a total or partial loss of function and the child present with a degree of morbidity.

4.7 CONCLUSIONS

Gene expression profiles have already been used to understand and even reverse dexamethasone resistance in ALL (Gullans, 2006), and continue to offer opportunities for new drug development (den Boer, 2007). The higher sensitivity of the latest 'chips' means that previously unseen LOH and copy number changes can be detected. Findings presented here for genomic copy number and loss of heterozygosity give a tantalising glimpse of what microarray analysis has to offer. Whilst direct evidence for mechanisms contributing to MTX resistance need to be followed up and proven, understanding allelic imbalance has the same potential that expression profiles have had in chemotherapy for ALL and other cancers. Reanalysis of this data once the long-term survival and outcome of each patient is known may provide essential information to more clearly categorise the patients and pinpoint the significance of individual changes in allelic imbalance.

5 CONCLUSIONS

Methotrexate has proven to be an effective chemotherapeutic agent in a number of malignancies, but because of its debilitating side effects and possible risk of developing secondary malignancy (Ebeo *et al.,* 2003; Gadner *et al.*, 2006), the oncologist must make an 'educated guess' at the minimum effective dose. Too high and the side-effects are intolerable, too low and the risk of developing resistance to this drug is increased with subsequent relapse. Mechanisms for methotrexate resistance include decreased uptake into the cell, principally involving the reduced folate carrier or enhanced excretion by the MRP, both reducing the intracellular concentration. Alternatively, increased expression or modified affinity of the target enzymes might influence the action of MTX as a competitive inhibitor.

Although the ideal approach would have been to compare the gDNA of children both at presentation and relapse, such paired material was very limited in the samples obtained under ethical approval for the current studies. The strategy used here to screen the gDNA at relapse was adopted because if mutations were to occur then they would be present at this stage and made the best use of the samples available. The weakness of this strategy would have been highlighted if mutations had been found in the relapse sample and there had not then been presentation material to determine if the mutation had existed before methotrexate exposure.

In spite of this potential flaw, evidence provided here supports the statement by Kaufman *et al* (2004), that "reduced folate carrier mutations are not the mechanism underlying methotrexate resistance in childhood acute lymphoblastic leukaemia". Somatic mutations in the coding sequence that could confer a loss of function, modify binding or facilitate selective uptake of methotrexate were not commonly found.

Although one patient who suffered multiple relapses had acquired a novel mutation in the 5'-UTR of the RFC-1 gene. Whilst not proven, the mutation may influence the transcription of the gene and/or stability of RFC mRNA, ultimately decreasing the amount of RFC available to transport MTX. Such down regulation has been reported in ALL and other cancers (Moscow *et* al., 1995; Gorlick *et al.*, 1997; Levy *et al.*, 2003; Rothem *et al.*, 2003; Wang *et al.*, 2003).

Similarly, screening of the DHFR gene did not provide any evidence to indicate that mutations in the coding sequence were contributing to MTX resistance. Both of these genes and the resulting proteins are critically important for cell growth, repair and replication. The findings here support the theory that any mutations are likely to be deleterious and account for their highly conserved nature. Therefore, the most likely mechanism for resistance is at the transcription and/or translation stages and would be influenced by copy number/-LOH and other epigenetic control mechanisms. In this study 'genechip' data was analysed to determine if such mechanisms were evident in these two genes and others involved in folate metabolism.

Since the introduction of the 10K microarray an understanding of the interrelationship between genes is becoming clearer, and yet, at the same time uncovering increasingly complexity. Genetic instability leading to malignancy was previously thought to occur at the megabase level, but recent studies using SNP microarrays have shown that increased copy number and LOH can occur across just a few kilobases (Zheng, 2005). Previously unseen gains could include intrachromosomal repeats or extrachromosomal double minutes resulting from chromosomal breaks (Singer *et al.*, 2000). LOH resulting from absolute loss of chromosomal material is intriguingly complicated by uniparental disomy so that tailoring chemotherapy to an individual

may be more difficult than at first thought, but even more necessary. The increased sensitivity offered by higher resolution 'chips' will allow more effective risk stratification and even personalised chemotherapy (Mullighan 2007; Teuffel 2004, Yeoh 2002). The data provided here does suggest that LOH in some patients may play a role in response to MTX, eg., loss of FPGH activity at presentation would mean that MTX polyglutamylated into its more active form will remain polyglutamylated for longer.

Published *in vivo* studies show that children with increased copy number of chromosome 21 have a better outcome, and it is this which appears to account for a good response to methotrexate, rather than a poor response due to resistance (Belkov *et al.*, 1999). This is supported here by microarray data for chromosome 21, showing that relapse patients have normal copy number and only the presentation group contains patients with higher estimated copy number. The scatter plot for FPGS (figure 4.15b) shows a similar distribution for estimated copy number, with presentation samples having apparent higher copy number which could provide them with higher enzyme activity and the capacity to polyglutamylate MTX into its more pharmacologically active form. Very few of the relapse samples have a higher estimated copy number and therefore are likely to maintain an activity expected in normal individuals and respond to MTX accordingly.

The knowledge base underpinning our understanding of cancer development, diagnosis and treatment has been growing exponentially for two decades and shows no signs of plateauing-off. Novel analytical techniques have already led to an understanding of the role that RFC down-regulation plays in MTX resistance allowing new antifolates to be developed that bypass RFC. Cell-lines cultured *in vitro* to be

MTX resistant were originally developed simply to study the molecular mechanisms of therapeutic failure. At the time it could not have been conceived that these cells could actually become part of the therapy itself. (Volpato *et al.,* 2007). The idea of transferring the mutated gene into the patient's normal stem cells and transfusing them back into the patient means that higher MTX can subsequently be given to kill the malignant clone and avoid severe side-effects to the normal cells.

5.1 Further work

One interesting finding of this study was that none of the relapse samples had the WT/WT (+19/+19) genotype in the 5'-UTR of the DHFR gene in ALL. Although the population sizes in this study were small, 73% of the relapse samples were scored as heterozygote WT/-19 whilst only 40% of the normal samples had this genotype. A loss of DHFR activity due to a loss of the proposed Sp1 transcription factor binding site (Xu *et al.,* 2007) could confer a greater degree of MTX resistance and needs to be investigated further on a larger sample group so that the difference can be statistically evaluated.

A set of primers and analytical techniques are available to rapidly genotype the VNTR status, either 2R/2R, 2R/3R or 3R/3R, of the TS gene. Looking at the distribution of copy number in figure 4.17b, the relapse group cluster around an estimated copy number of 2.0, whilst the presentation samples group at 2.5-3.0. If the gDNA is available on these samples then it would be useful to use the method developed to determine if there is any association between the VNTR status and outcome (Krajinovic *et al.,* 2002).

The significance of the novel mutation found in the 5'-UTR of the RFC-1 gene in sample 6250 should be followed up by measuring transcription rates. Real-time reverse transcription PCR will indicate if this gene has been down regulated as a result of the mutation. Analysis of a larger number of relapse childhood ALL samples will provide information on the occurrence of this mutation and its contribution to MTX resistance.

As the gDNA pools are 'expanded' and DNA becomes available then several of the initial findings can be repeated to confirm the results or followed-up by sequencing, especially DHFR exon 4 and exon 5 heteroduplex patterns. If cell-lines with mutations or novel SNPs in the amplified regions of RFC exon 5 and DHFR exons 3 and 5 are published, then these can be used as positive controls to confirm that DHPLC conditions are optimal for picking up heteroduplexes and thus meet the Diagnostic DHPLC Quality Assurance criteria (Schollen *et al.*, 2005).

6 REFERENCES

Agarwal P K, Perl A. PCR amplification of highly GC-rich DNA template after denaturation by NaOH. *Nucleic Acids Research* 1993; **21**: 5283-884.

Agarwal P K, Billeter S R, Rajagopalan P T R, Benkovic S J, Hammes-Schiffer S. Network of coupled promoting motions in enzyme catalysis. *Proceedings of the National Academy of Science* 2002; **99**: 2794-2799.

Anagnou N P, O'Brien S J, Shimada T, Nash W G, Chen M-J, Nienhuis A W. Chromosomal organization of the human dihydrofolate reductase genes: dispersion, selective amplification, and a novel form of polymorphism. *Proceedings of the National Academy of Science* 1984; **81**: 5170-5174.

Anon. GeneChip® Human Mapping 100k set. Affymetrix UK Ltd, High Wycombe. 2004.

Anon. High-resolution chromosome copy number analysis using GeneChip® mapping arrays. Affymetrix UK Ltd, High Wycombe. 2005

Anon. Segmented neutrophil. *http://www.wadsworth.org/chemheme/heme/microscope/seg.htm* Visited 28/1/2006a.

Anon. Mature lymphocyte. *http://www-medlib.med.utah.edu/WebPath/HEMEHTML/HEME002.html* Visited 28/1/2006b.

Anon. GeneCard for protein-coding MTR *http://www.genecards.org/cgi-bin/carddisp.pl?gene=MTR&search=methionine+synthase* Visited 27/5/2006c.

Arico M, Valsecchi MG, Camitta B, Schrappe M, Chessells J, Baruchel A, Gaynon P, Silverman L, Janka-Schaub G, Kamps W, Pui CH, Masera G. Outcome of treatment in children with Philadelphia chromosome-positive acute lymphoblastic leukaemia. *New England Journal of Medicine* 2000; **342** :998-1006.

Asai S, Miyachi H, Kobayashi H, Takemura Y, Ando Y. Large diversity in transport-mediated methotrexate resistance in human leukaemia cell line CCRF-CEM established in a high concentration of leucoverin. *Cancer Science* 2003; **94**: 210-214.

Asakura K, Uchida H, Miyachi H, Kobayashi H, Miyakawa Y, Nimer S D, Takahashi H, Ikeda Y, Kizaki M. TEL/AML1 overcomes drug resistance through transcripitional repression of multidrug resistance-1 gene expression. *Molecular Cancer Research* 2004; **2**: 339-347.

Assaraf Y G. Molecular basis of antifolate resistance. *Cancer Metastasis Reviews* 2007: **26**: 153-181.

Avery L. Mann-Whitney U test. University of Texas Southwestern Medical Center *http://elegans.swmed.edu/~leon/stats/utest.html Visited 24/11/2007.*

Bahrami A R, Dickman M J, Matin M M, Ashby J R, Brown P E, Controy M J, Fowler G J S, Rose J P, Sheikh Q I, Yeung A T, Hornby D P. Use of fluorescent DNA-intercalating dyes in the analysis of DNA via ion-pair reversed-phase denaturing high-performance liquid chromatography. *Analytical Biochemistry* 2002; **309**: 248-252.

Banerjee D, Mayer-Kuckuk P, Cpiaux G, Budak-Alpdogan T, Gorlick R, Bertino J R. Novel aspects of resistance to drugs targeted to dihydrofolate reductase and thymidylate synthase. *Biochimica et Biophysica Acta* 2002; **1587**: 164-173.

Barrett J C, Kawasaki E S. Microarrays: the use of the oligonucleotides and cDNA for the analysis of gene expression. *Drug Discovery Today* 2003; **8**: 134-141.

Bayat A, Walter J, Lamb H, Marino M, Ferguson M W J, Ollier W E R. Mitochondrial mutation detection using enhanced multiplex denaturing high-performance liquid chromatography. *International Journal of Immunogenetics* 2005; **32**: 199-205.

Beard W A, Appleman J R, Huang S, Delcamp T J, Freisheim J H, Blakley R L. Role of the conserved active site residue tryptophan-24 of human dihydrofolate reductase as revealed by mutagenesis. *Biochemistry* 1991; **30**: 1432-1440.

Belkov V M, Krynetski E Y, Schuetx J D, Yanishevski Y, Masson E, Mathew S, Raimondi S, Pui C-H, Relling M V, Evans W E. Reduced folate carrier expression in acute lymphoblastic leukemia: A mechanism for ploidy but not lineage differences in methotrexate accumulation. *Blood* 1999; **93**: 1643-1650.

Belur L R, Boelk-Galvan D, Diers M D, McIvor R S, Zimmerman C L. Methotrexate accumulates to similar levels in animals transplanted with normal versus drug-resistant transgenic marrow. *Cancer Research* 2001; **61**: 1522-1526.

Bennet J M, Catovsky D, Daniel M T, Flondrin G, Galton D A G, Gralnick H R, Sultan C. The morphological classification of acute lymphoblastic leukaemia: concordance among observers and clinical correlations. *British Journal of Haematology* 1981; **47**: 553-61.

Beroukhim R, Lin M, Park Y, Hao K, Zhao X, Garraway L A, Fox E A, Hochberg E P, Mellinghoff I K, Hofer M D, Descazeaud A, Rubin M A, Meyerson M, Wong W H, Sellers W R, Li C. Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Computational Biology* 2006; **2**: e41. DOI: 10.1371/journal.pcbi.0020041.

Bertino J R, Donohue D M, Dimmons B, Gabrio B W, Silber R, Huennekens F M. The "induction" of dihydrofolic reductase activity in leukocytes and erythrocytes of patients treated with amethopterin. *Journal of Clinical Investigation* 1963; **42**: 466-475.

Bisseling T M, Steegers E A, van den Heuvel J J, Siero H L, van de Water F M, Walker A J, Steegers-Theunissen R P, Smits P, Russel F G. Placental folate transport and binding are not impaired in pregnancies complicated by fetal growth restriction. *Placenta* 2004; **25**: 588-593.

Blakley R L, Appleman J R, Chunduru S K, Nakano T, Lewis W S, Harris S E. Mutations of human dihydrofolate reductase causing decreased inhibition by methotrexate. *Advances in Experimental Medical Biology* 1993; **338**: 473-479.

Blakley R L, Sorrentino B P. In-vitro mutations in dihydrofolate reductase that confer resistance to methotrexate: potential for clinical application. *Human Mutation* 1998; **11**: 259-263.

Blount B C, Mack M M, Wehr C M, MacGregor J T, Hiatt R A, Wang G, Wickramasinge S N, Everson R B, Ames B N. Folate deficiency causes uracil misincorporation into human and chromosome breakage: Implications for cancer and neuronal damage. *Proceedings of the National Academy of Science* 1997; **94**: 3290-3295.

Bosson G. Reduced folate carrier: Biochemistry and molecular biology of the normal and methotrexate-resistant cell. *British Journal of Biomedical Science* 2003; **60**: 117-129.

Brigle K E, Spinella M J, Westin E H, Goldman I D. Increased expression and characterization of two distinct folate binding proteins in murine erythroleukemia cells. *Biochemical Pharmacology* 1994; **47**: 337-345.

Brigle K E, Spinella M J, Sierra E E, Goldman I D. Characterization of a mutations in the reduced folate carrier in a transport defective L1210 murine leukemia cell line. *Journal of Biological Chemistry* 1995; **270**: 22974-22979.

Brigle K E, Spinella M J, Sierra E E, Goldman I D. Organization of the murine reduced folate carrier gene and identification of variant splice forms. *Biochemica et Biophysica Acta* 1997; **1353**: 191-198.

Brodsky G, Barnes T, Bleskan J, Becker L, Cox M, Patterson D. The human GARS-AIRS-GART gene encodes two proteins which are differentially expressed during human brain development and temporally overexpressed in cerebellum of individuals with Down syndrome. *Human Molecular Genetics* 1997; **6**: 2043-2050.

Bullerjahn A M E, Freisheim J H. Site-directed deletion mutants of a carboxyl-terminal region of human dihydrofolate reductase. *Journal of Biological Chemistry* 1992; **267**: 864-870.

Butcher L M, Davis O S P, Craig I W, Plomin R. Genome-wide quantitative trait locus association scan of genera cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes, Brain and Behaviour* 2007; doi: 10.1111/j.1601-183X.2007.00368. Available at *www.blackwell-synergy.com*

Cadieux B, Ching T-T, Vandenberg S R, Costello J F. Genome-wide hypoemthylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation. *Cancer Research* 2006; **66**: 8469-8476.

Campbell P N, Smith A D, Peters T J. *Biochemistry Illustrated* (5th Edition) Elsevier Churchill Livingstone: Edinburgh. 2005: 104.

Cancer Genome Anatomy Project. SNP500 Cancer Database; **rs12659**. National Cancer Institute. *http://snp500cancer.nci.nih.gov/snp.cfm?both_snp_id=SLC19A1-02*. visited 16/12/2006.

Cancer Genome Anatomy Project. SNP500 Cancer Database; **rs1051269**. National Cancer Institute. *http://snp500cancer.nci.nih.gov/snp.cfm?both_snp_id=SLC19A1-02*. visited 3/2/2007.

Cancer Research UK. Incidence – Leukaemia. *http://info.cancerresearchuk.org/cancerstats/types/leukaemia/incidence/* Visited 12/11/2005

Cao W, Matherly L H. Analysis of the membrane topology for transmembrane domains 7-12 of the human reduced folate carrier by scanning cysteine accessibility methods. *Biochemical Journal* 2004; **378**: 201-206.

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane C R, Lim E P, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutx R, Daley G Q, Lander E S. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics* 1999; **22**: 231-238.

Carotta S, Nutt S L. Losing B cell identity. *BioEssays* 2008; **30**: 203-207.

Carroll W L, Bhojwani D, Min D-J, Raetz E, Relling M, Davies S, Downing J R, Willman C L, Reed J C. Pediatric acute lymphoblastic leukaemia. *Hematology (Am Soc Hematol Educ Program)* 2003; 102-31.

Carroll W L, Bhojwani D, Min D-J, Moskowitz N, Raetz E A. Childhood acute lymphoblastic leukaemia in the age of genomics. *Pediatric Blood Cancer* 2006; **46**: 570-578.

Chakrabarti R, Schutt C E. The enhancement of PCR amplification by low molecular weight amides. *Nucleic Acids Research* 2001a; **29**: 2377-2381.

Chakrabarti R, Schutt C E. The enhancement of PCR amplication by low molecular weight sulfones. *Gene* 2001b; **274**: 293-298.

Chango A, Emery-Fillon N, Potier de Courcy G, Lambert D, Pfistser M, Rosenblatt D S, Nicolas J-P. A polymoprhism (80G-A) in the reduced folate carrier gene and its association with folate status and homocycteinema. *Molecular Genetics and Metabolism* 2000; **70**: 310-315.

Chango A, Willequet F, Fillon-Emery N, Nicholas J P, Blehaut H. The single nucleotide polymorphism (80G→A) of reduced folate carrier gene in trisomy 21. *American Journal of Clinical Nutrition* 2004; **80**: 1667-1669.

Chave K J, Ryan T J, Chmura S E, Galivan J. Identification of single nucleotide polymorphisms in the human γ-glutamyl hydrolase gene and characterization of promoter polymorphisms. *Gene* 2003; **319**: 167-175.

Chen L, Qi J. Korenberg J, Garrow T A, Choi Y-J, Shane B. Purification and properties of human cytosolic polypoly-γ-glutamate synthetase and organization, localization, and differntial splicing of its gene. *Journal of Biological Chemistry* 1996; **271**: 13077-13087.

Chen M-J, Shimada T, Moulton A D, Harrison M, Nienhuis A W. Intronless human dihydrofolate reductase genes are derived from processed RNA molecules. *Proceedings of the National Academy of Science* 1982; **79**: 7435-7439.

Chen M-J, Shimada T, Moulton A D, Cline A, Humphries R K, Maizel J, Nienhuis A W. The functional human dihydrofolate reductase gene. *Journal of Biological Chemistry* 1984; **259**: 3933-3943.

Chen Z-S, Lee K, Walther S, Roftogianis R B, Kuwano M, Zheng H, Kruh G D. Analysis of methotrexate and folate transport by multidrug resistance protein 4 (ABCC4): MRP4 is a component of the methotrexate efflux system. *Cancer Research* 2002; **62**: 3144-3150.

Cheng Q, Wu B, Kager L, Panetta J C, Zheng J, Pui C-H, Relling M V, Evans W E. A substrate specific functional polymorphism of human γ-glutamyl hydrolase alters catalytic activity and methotrexate polyglutamate accumulation in acute lymphoblastic leukaemia cells. *Pharmacogenetics* 2004; **14**: 557-567.

Cheng Q, Cheng C, Crews K R, Ribeiro R C, Pui C-H, Relling M V, Evans W E. Epigenetic regulation of human γ-glutamyl hydrolase activity in acute lymphoblastic leukemia cells. *American Journal of Human Genetics* 2006; **79**: 264-274.

Cheng S, Fockler C, Barnes W M, Higuchi R. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proceedings of the National Academy of Science* 1994; **91**: 5695-5699.

Cheok M H, Evans W E. Acute lymphoblastic leukaemia: a model for the pharmacogenomics of cancer therapy. *Nature Reviews Cancer* 2006; **6**: 117-129.

Chessells J M, Bailey C, Richards S M. Intensification of treatment an survival in all children with lymphoblastic leukaemia: results of UK Medical Research Council trial UKALL X. *The Lancet* 1995; **345**: 143-148.

Chessells J M, Harrison G, Richards S M, Bailey C C, Hill F G H, Gibson B E, Hann I M. Down's syndrome and acute lymphoblastic leukaemia: clinical features and response to treatment. *Archives of Disease in Childhood* 2001; **85**: 321-325.

Chiaretti S, Tavolaro S, Ghia E M, Ariola C, Metteucci C, Elia L, Maggio R, Messina M, Ricciardi M R, Vitale A, Ritz J, Mecucci C, Guarini A, Foa R. Charecterization of *ABL1* expression in adult T-cell acute lymphoblastic leukenmia by oligonucleotide array analysis. *Haematologica* 2007; **92**: 619-626.

Chou L-S, Lyon E, Wittwer C T. A comparison of high-resolution melting analysis with denaturing high-performance liquid chromatography for mutation scanning. *American Journal of Clinical Pathology* 2005; **124**: 330-338.

Choudhuri S. Microarrays in biology and medicine. *Journal of Biochemical and Molecular Toxicology* 2004; **18**: 171-179.

Chu E, Drake J C, Boarman D, Baram J, Allegra C J. Mechanism of thymidylate synthase inhibition by methotrexate in human neoplastic cell lines and normal human myeloid progenitor cells. *Journal of Biological Chemistry* 1990; **265**: 8470-8478.

Chu E, Koeller D M, Casey J L, Drake J C, Chabner B A, Elwood P C, Zinn S, Allegra C J. Autoregulation of human thymidylate synthase messenger RNA translation by thymidylate synthase. *Proceedings of the National Academy of Science* 1991; **88**: 8977-8981.

Chu E, Takimoto C H, Voeller D, Grem J L, Allegra C J. Specific binding of human dihydrofolate reductase protein in dihydrofolate reductase messenger RNA in vitro. *Biochemistry* 1993; **32**: 4756-4760.

Chua S-W, Vijayakumar P, Nissom P M, Yam C-Y, Wong V V T, Yang H. A novel normalization method for effective removal of systematic variation in microarray data. *Nucleic Acids Research* 2006; **34**: e38.

Chunduru S K, Cody V, Luft J R, Pangborn W, Appleman J R, Blakley R L. Methotrexate-resistant variants of human dihydrofolate reductase: effects of Phe^{31} substitutions. *Journal of Biological Chemistry* 1994; **269**: 9547-9555.

Cody V, Luft J R, Pangborn W. Understanding the role of Leu22 variants in methotrexate resistance: comparison of wild-type and Leu22Arg variant mouse and human dihydrofolate reductase ternary crystal complexes with methotrexate and NADPH. *Acta Crystallographica* 2005; **D61**: 147-155.

Colley J, Jones S, Dallosso A R, Maynard J H, Humphreys V, Dolwani S, Sampson J R, Cheadle J P. Rapid recognition of aberrant dHPLC elution profiles using the Transgenomic Navigator™ software. *Human Mutation* 2005; **26**: 165-173.

Cotton R G H. Slowly but surely towards better scanning for mutations. *Trends in Genetics* 1997; **13**: 43-46.

Cotton R G H, Bray P J. Using CCM and DHPLC to detect mutations in the glucocorticoid receptor in atherosclerosis: a comparison. *Journal of Biochemical and Biophysical Methods* 2001; **47:** 90-100.

Cowan K H, Goldsmith M E, Levine R M, Aitken S C, Douglass E, Clendeninn N, Nienhuis A W, Lippman M E. Dihydrofolate reducatase gene amplification and possible rearrangement in estrogen-responsive methotrexate-resistant human breast cancer cells. *Journal of Biological Chemistry* 1982; **257**: 15079-15086.

Cui X, Churchill G A. Statistical tests for differential expression in cDNA microarray experiments. *BioMed Central Genome Biology* 2003; **4**: 210.
Available at *http://genomebiology.com/2003/4/4/210*

Dalma-Weiszhausz D D, Murphy G M. Single nucleotide polymorphisms and their characterization with oligonucleotide microarrays. *Psychiatric Genetics* 2002; **12**: 97-107.

Daniel S, Mytelka M J C. Analysis and suppression of DNA polymerase pauses associate with a trinucletide consensus. *Nucleic Acids Research* 1996; **24**: 2774-81.

Daubner S C, Young M, Sammons R D, Courtney L F, Benkovic S J. Structural and mechanistic studies on the HeLa anc hicken liver proteins that catalyse glycinamide ribonucleotide synthesis and formylation and aminoimidazole ribonucleotide synthesis. *Biochemistry* 1986; **25**: 2951-2957.

Davies S M, Borowitz M J, Rosner G L, Ritz K, Devidas M, Winick N, Martin P L, Bowman P, Elliott J, Willman C, Das S, Cook E H, Relling M V. Pharmacogenetics of minimal residual disease response in children with B-precursor acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood* 2008; **111**: 2984-2990.

De Jonge R. Clinical consequences of polymorphisms in the methylenetetrahydrofolate reductase gene. **Programme & Abstracts, Joint 11th International and 9th European Symposium on Purines and Pyrimidines in man:** 2003; PP03: 041.

De Jonge R, Hooijberg J H, van Zelst, Jansen G, van Zantwijk C H, Kaspers G J L, Peters F G J, Ravindranath Y, Pieters R, Lindemans J. Effect of polymorphisms in folate-related genes on in-vitro methotrexate sensitivity in pediatric acute lymphoblastic leukemia. *Blood* 2005; **106**: 717-720.

De Lathouder S, Gerards A H, de Groot E R, Valkhof M G, Dijkmans B A, Aarden L A. Bioassay for detection of methotrexate in serum. *Scandinavian Journal of Rheumatology* 2004; **33**: 167-173.

De Marco P, Calevo M G, Moroni A, Arata L, Merello E, Carna A, Finnell R H, Andreussi L, Capra V. Polymorphisms in genes involved in folate metabolism as risk factors for NTDs. *European Journal of Pediatric Surgery* 2001; **11 suppl 1**: S14-S17.

Dehainault C, Lauge A, Caux-Moncoutier V, Pages-Berhouet S, Doz F, Desjardins L, Couturier J, Gauthier-Villars M, Stoppa-Lyonnet D, Houdayer C. Multiplex PCR/liquid chromatography assay for full detection of gene rearrangements: application to RB1 gene. *Nucleic Acids Research* 2004; **32**: e139.

Den Boer M L, Pieters R. Microarray-based identification of new targets for specific therapies in pediatric leukemia. *Current Drug Targets* 2007; **8**: 761-764.

Dervieux T, Furst D, Lein D O, Capps R, Smith K, Walsh M, Kremer J. Polyglutamation of methotrexate with common polymorphisms in reduced folate carrier, aminoimidazole carboxamide ribonucleotide transformylase, and thymidylate synthetase are associated with methotrexate effects in rheumatoid arthritis. *Arthritis and Rheumatism* 2004a; **50**: 276-2774.

Dervieux T, Kremer J, Lein D O, Capps R, Barham R, Meyer G, Smith K, Caldwell J, Furst D E. Contribution of common polymorphisms in reduced folate carrier and γ-glutamylhydrolase to methotrexate polyglutamate levels in patients with rheumatoid arthritis. *Pharmacogenetics* 2004b; **14**: 733-739.

Dicker A P, Volkenandt M, Bertino J R. Detection of a single base mutation in the human dihydrofolate reductase gene from a methotrexate-resistant cell line using the polymerase chain reaction. *Cancer Communications* 1989; **1**: 7-12.

Dicker A P, Waltham M C, Volkenandt M, Schweitzer B I, Otter G M, Schmid F A, Sirotnak F N, Bertino J R. Methotrexate resistance in an *in vivo* mouse tumour due to a non-active-site dihydrofolate reductase mutation. *Proceedings of the National Academy of Science* 1993; **90**: 11797-11801.

Digiuseppe J A. Acute lymphoblastic leukaemia: diagnosis and detection of minimal residual disease following therapy. *Clinical Laboratory Medicine* 2007; **27**: 533-549.

Ding B C, Whetstine J R, Witts T L, Schuetz J D, Matherly L H. Repression of human reduced folate carrier gene expression by wild type p53. *Journal of Biological Chemistry* 2001a; **276**: 8713-8719.

Ding B C, Witt T L, Hukku B, Heng H, Zhang L, Matherly L H. Association of deletions and translocation of the reduced foalte carrier gene with profound loss of gene expression in methotrexate-resistant K562 human erythroleukemia cells. *Biochemical Pharmacology* 2001b; **61**: 665-675.

Dixon K H, Mulligan T, Chung K-N, Elwood P C, Cowan K H. Effects of folate receptor expression following stable transfection into wild type and methotrexate transport-deficient ZR-75-1 human breast cancer cells *Journal of Biological Chemistry* 1992; **267**: 24140-24147.

Dixon K H, Lampher B C, Chiu J, Kelly K, Cowan K H. A novel cDNA restores reduced folate carrier activity and methotrexate sensitivity to transport deficient cells. *Journal of Biological Chemistry* 1994; **269**: 17-20.

Dobbin K K, Beer D G, Meyerson M, Yeatman T J, Gerald W L, Jacobson J W, Conley B, Buetow K H, Heiskanen M, Simon R M, Minna J D, girard L, Misek D E, Taylro J M G, Hanash S, Naoki K, Hayes N, Ladd-Acosta C, Enkemann S A, Viale A, Giordano T J. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clinical Cancer Research* 2005; **11**: 565-572.

Don R H, Cox P T, Wainwright B J, Baker K, Mattick J S. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* 1991; **19**: 4008.

Donadieu J, Hill C. Early response to chemotherapy as a prognostic factor in childhood acute lymphoblastic leukaemia: a methodological review. *British Journal of Haematology* 2001; **115**: 34-45.

Doty P, Marmur J, Eigner J, Schildkraut C. Strand separation and specific recombination in deoxyribonucleic acids: physical chemical studies. *Biochemistry* 1960; **46**: 461-476.

Drori S, Sprecher H, Shemer G, Jansen G, Goldman I D, Assaraf Y G. Characterization of a human alternatively spliced truncated reduced folate carrier increasing folate accumulation in parental leukemia cells. *European Journal of Biochemistry* 2000a; **267**: 690-702.

Drori S, Jansen G, Mauritz R, Peters G J, Assaraf Y G. Clustering of mutations in the first transmembrane domain of the human reduced folate carrier in GW1843U89-resistant leukemia cells with impaired antifolate transport and augmented folate uptake. *Journal of Biological Chemistry* 2000b; **275**: 30855-30863.

Dulucq S, St-Onge G, gagne V, Ansari M, Sinnett D, Labuda D, Mogharabi A, Kranjinovic M. DNA variants in the dihydrofolate reductase gene and outcome in childhood ALL. *Blood* 2008; **111**: 3692-3700.

Dunphy C H. Gene expression profiling data in lymphoma and leukaemia. *Archives of Pathology and Laboratory Medicine* 2006; **130**: 483-520.

Duthie S J. Folic acid deficiency and cancer: mechanisms of DNA instability. *British Medical Bulletin* 1999; **55**: 578-592.

Eastman H B. Swick A G, Schmitt M C, Azizkhan J C. Stimulation of dihydrofolate reductase promoter activity by antimetabolic drugs. *Proceedings of the National Academy of Science* 1991; **88**: 8572-8576.

Ebeo C T, Girish M R, Byrd R P, Roy T M, Mehta J B. Methotrexate-induced pulmonary lymphoma. *Chest* 2003; **123**: 2150-2153.

Eisen M B, Spellman P T, Brown O P, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Science* 1998; **95**: 14863-14868.

Engel E. Uniparental disomies I unselected populations. *Americal Journal of Human Genetics* 1998; **63**: 962-966.

Ercikan-Abali E A, Wlatham M C, Dicker A P, Schweitzer B I, Gritsman H, Banerjee D, Bertino J R. Variants of human dihydrofolate reductase with substitutions at leucine-22: effects on catalytic and inhibitor binding properties. *Molecular Pharmacology* 1996; **49**: 430-437.

Ezzeldin H, Okamoto Y, Johnson M R, Diasio R B. A high-throughput denaturing high-performance liquid chromatography method for the identification of variant alleles associated with dihydropyrimidine dehydrogenase deficiency. *Analytical Biochemistry* 2002; **306**: 63-73.

Ezzeldin H, Hoffmayer C, Soong R, Johnson M R, Lee A, Heslin M, Diasio R. Simultaneous detection of variable number tandem repeats, single nucleotide polymorphisms, and allelic imbalance in the thymidylate synthase gene enhancer region using denaturing high-performance liquid chromatography. *Analytical Biochemistry* 2004; **334**: 276-283.

Faderl S, Jeha S, Kantarjian H M. The biology and therapy of adult acute lymphoblastic leukaemia. *Cancer* 2003; **98**: 1337-54.

Farber S, Diamond L K, Mercer R D, Sylvester R F Jr, Wolff J A. Temporary remission in acute leukemia in children produced by folic acid antagonist, 4-aminopteryl;-glutamic acid (aminopterin). *New England Journal of Medicine* 1948; **238**: 787-793.

Ferguson P L, Flintoff W F. Topological and functional analysis of the human reduced folate carrier by hemaglutinin epitope insertion. *Journal of Biological Chemistry* 1999; **274**: 16269-16278.

Ferl G Z, Timmerman J M, Witte O N. Extending the utility of gene profiling data by bridging microarray platforms. *Proceedings of the National Academy of Science* 2003; **100**: 10585-10587.

Fillon-Emery N, Chango A, Mircher C, Barbe F, Belhaut H, Herbeth B, Rosenblatt D S, Rethore M-O, Lambert D, Nicholas J P. Homocysteine concentrations in adults with trisomy 21: effect of B vitamins and genetic polymorphisms. *American Journal of Clinical Nutrition* 2004; **80**: 1551-1557.

Findley H W, Gu L, Yeager A M, Zhou M. Expression and regulation of Bcl-2, Bcl-xl, and Bax correlate with p53 status and sensitivity to apoptosis in childhood acute lymphoblastic leukemia. *Blood* 1997; **89**: 2986-2993.

Flatley R M, Payton S G, Taub J W, Matherly L H. Primary acute lymphoblastic leukaemia cells use a novel promoter and 5'noncoding exon for the human reduced folate carrier that encodes a modified carrier translated from an upstream translational start. *Clinical Cancer Research* 2004; **10**: 5111-5122.

Flintoff W F, Williams F M R, Sadlish H. The region between transmembrane domains 1 and 2 of the reduced folate carrier forms part of the substrate-binding pocket. *Journal of Biological Chemistry* 2003; **278**: 40867-40876.

Flintoff W F, Sadlish H, Gorlick R, Yang R, Williams F M R. Functional analysis of altered reduced folate carrier sequence changes identified in osteosarcomas. *Biochimica et Biophysica Acta* 2004; **1690**: 110-117.

Fodor S P A, Read J L, Pirrung M C, Stryer L, Lu T, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science* 1991; **251**: 767-773.

Frackman *et al*, Kobs G, Simpson D, Storts D. Betaine and DMSO: Enhancing agents for PCR. Promega Corporation, Madison, Wisconsin. *http://www.promega.com/pnotes/65/6971_27/dfault.html* visited 12/10/2002.

Friso S, Choi S-W, Girelli D, Mason J B, Dolnikowski G G, Bagley P J, Olivieri O, Jacques P F, Rosenberg I H, Corrocher R, Selhub J. A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status. *Proceedings of the National Academy of Science* 2002; **99**: 5606-5611.

Frueh F W, Noyer-Weidner M. The use of denaturing high-performance liquid chromatography (DHPLC) for the analysis of genetic variations: impact for diagnostics and pharmacogenetics. *Clinical Chemistry Laboratory Medicine* 2003; **41**: 452-461.

Funanage V L, Myoda T T, Moses P A, Cowell H R. Assignment of the human dihydrofolate reductase gene to the q11→q22 region of chromosome 5. *Molecular and Cellular Biology* 1984; **4**: 2010-2016.

Gadner H, Masera G, Schrappe M, Eden E, Benoit Y, Harrison C, Ncahman J, Pui C-H. The eighth international childhood acute lymphoblastic leukemai workshop ('Pointe di Legno meeting') report: Vienna, Austria, April 27-28, 2005. *Leukemia* 2006; **20**: 9-17.

Galivan J, Ryan T, Rhee M, Yao R, Chave K. Glutamyl hydrolase: properties and phamacologic impact. *Seminars in Oncology* 1999; **26**: 33-37.

Galivan J, Ryan T J, Chave K, Rhee M, Yao R, Yin D. Glutamyl hydrolase: pharmacological role and enzymatic characterization. *Pharmacology and Therapeutics* 2000; **85**: 207-215.

Gifford A J, Kavallaris M, Madafiglio J, Matherly L H, Stewart B W, Haber M, Norris M D. P-glycoprotein-mediated methotrexate resistance in CCRF-CEM sublines deficient in methotrexate accumulation due to a point mutation in the reduced folate carrier gene. *International Journal of Cancer* 1998; **78**: 176-181.

Gifford A J, Haber M, Witt T L, Whetstine J R, Taub J W, Matherly L H, Norris M D. Role of the E45K-reduced folate carrier gene mutation in methotrexate resistance in human leukemia cells. *Leukemia* 2002; **16**: 2379-2387.

Giordano M, Oefner P J, Underhill P A, Luca C S L, Tosi R and Monigliano R P. Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. *Genomics* 1999: **56**; 247-253.

Giordano M, Mellai M, Hoogendoorn B, Momigliano-Richiardi P. Determination of SNP allele frequencies in pooled DNAs by primer extension genotyping and denaturing high-performance liquid chromatography. *Journal of Biochemical and Biophysical Methods* 2001; **47**: 101-110.

Goker E, Waltham M, Kheradpour A, Trippett T, Mazumdar M, Elisseyeff Y, Schnieders B, Steinherz P, Tan C, Berman E, Bertino J R. Amplification of the dihydrofolate reductase gene is a mechanism of acquired resistance to methotrexate in patients with acute lymphoblastic leukemia and is corrected with p53 gene mutations. *Blood* 1995; **86**: 677-684.

Goldman I D. The characteristics of the membrane transport of amethopterin and the naturally occurring folates. *Annals of the New York Academy of Science* 1971; **186**: 400-422.

Gong M, Yess J, Connolly T, Ivy S P, Ohnuma T, Cowan K H, Moscow J A. Molecular mechanism of antifolate transporter-deficiency in a methotrexate-resistant MOLT-3 human leukemia cell line. *Blood* 1997; **89**: 2494-2499.

Gong M, Cowan K H, Gudas J, Moscow J A. Isolation and characterization of genomic sequence involved in the regulation of the human reduced folate carrier gene (RFC1). *Gene* 1999; **233**: 21-31.

Gorlick R, Goker E, Trippett T, Wlatham M, Banerjee D, Betino J R. Intrinsic and acquired resistance to methotrexate in acute leukemia. *New England Journal of Medicine* 1996; **335**: 1041-1048.

Gorlick R, Goker E, Trippett T, Steinherz P, Elisseyeff Y, Mazumdar M, Flintoff W F, Bertino J R. Defective transport is a common mechanism of acquired methotrexate resistance in acute lymphocytic leukemia and is associated with decreased reduced folate carrier expression. *Blood* 1997; **89**: 1013-1018.

Goto Y, Yue L, Yokoi A, Nishimura R, Uehara T, Koizumi S, Saikawa Y. A novel single-nucleotide polymorphism in the 3'-untranslated region of the human dihydrofolate reductase gene with enhanced expression. *Clinical Cancer Research* 2001; **7**: 1952-1956.

Gottardo N G, Hoffmann K, Beesley A H, Freitas J R, Firth M J, Perera K U, de Klerk N H, Baker D L, Kees U R. Identification of novel molecular prognostic markers for paediatric T-cell acute lymphoblastic leukaemia. *British Journal of Haematology* 2007; **137**: 319-328.

Goulian M, Bleile B, Tseng B Y. Methotrexate-induced misincorporation of uracil into DNA. *Proceedings of the National Academy of Science* 1980; **77**: 1956-1960.

Goyette P, Sumner JS, Milos R, Duncan AMV, Rosenblatt DS, Matthews RG, Rozen R. Human methylenetetrahydrofolate reductase: isolation of cDNA, mapping and mutation identification. *Nature Genetics* 1994; **7**: 195-200.

Goyette P, Pai A, Milos R, Frosst P, Tran P, Chen Z, Chan M, Rozen R. Gene structure of human and mouse methylenetetrahdyrofolate reductase (MTHFR). *Mammalian Genome* 1998; **9**: 652-656.

Greaves M. In utero origins of childhood leukaemia. *Early Human Development* 2005; **81**: 123-129.

Gross E, Arnold N, Pfeifer K, Bandick K, Kiechle M. Identification of specific BRCA1 and BRCA2 variants by DHPLC. *Human Mutation* 2000; **16**; 345-353.

Gross E, Kiechle M, Arnold N. Mutation analysis of p53 in ovarian tumors by DHPLC. *Journal of Biochemical and Biophysical Methods* 2001; **47**: 73-81.

Gullans S R. Connecting the dots using gene-expression profiles. *New England Journal of Medicine* 2006; **355**: 2042-2043.

Guo W, Healey J H, Meyers P A, Ladanyi M, Guvos A G, Bertino J R, Gorlick R. Mechanisms of methotrexate resistance in osteosarcoma. *Clinical Cancer Research* 1999; **5**: 621-627.

Hamel E, Johnson G, Glubiger D. Pharmacokinetics of leucoverin rescue using a new methotrexate-independent biochemical assay for leucoverin and N^5-methyltetrahydrofolate. *Cancer Treatment Reports* 1981; **65**: 545-553.

Hanauer D A, Rhodes D R, Sinha-Kumar C, Chinnaiyan A M. Bioinformatics approaches in the study of cancer. *Current Molecular Medicine* 2007; **7**: 133-141.

Hann I, Vora A, Harrison G, Eden O, Hill F, Gibson B, Richards S. Determinants of outcome after intensified therapy of childhood lymphoblastic leukaemia: results from Medical Research Council United Kingdon acute lymphoblasstic leukaemia XI protocol. *British Journal of Haematology* 2001; **113**: 103-114.

Hannun Y A. Apoptosis and the dilemma of cancer chemotherapy. *Blood* 1997; **89**: 1845-1853.

Harrison C J. The detection and significance of chromosomal abnormalities in childhood acute lymphoblastic leukaemia. *Blood Reviews* 2001; **15**: 49-59.

Harrison C J, Griffiths M, Moorman F, Schnittger S, Cayuela J-M, Shurtleff S, Gottardi E, Mitterbauer G, Colmer D, Delabesse E, Casteras V. A multicenter evaluation of comprehensive analysis of MLL translocations and fusion gene partners in acute leukaemia using the MLL FusionChip device. *Cancer Genetics and Cytogenetics* 2007; **173**: 17-22.

Hart A. Mann-Whitney test is not just a test of medians: differences in spread can be important. *British Medical Journal* 2001; **323**: 391-393.

Hawkins D. *Biomeasurement.* Oxford University Press, Oxford. 2005, 142-152.

Heerema N A, Nachman J B, Sather H N, Sensel M G, Lee M K, Hutchinson R, Lange B J, Steinherz P G, Bostrom B, Gaynon P S, Uckun F. Hypodiploidy with less than 45 chromosomes confers adverse risk in childhood acute lymphoblastic leukemia: A report from the Children's Cancer Group. *Blood* 1999; **94**: 4036-4046.

Heil S G, Lievers K J A, Boers G H, Verhoef P, den Heijer M, Trijbels F J M, Blom H J. Betaine-methyltransferase (BHMT): genomic sequencing and relevance to hyperhomocysteineami and vascular disease in humans. *Molecular Genetics and Metabolism* 2000; **71**: 511-519.

Hengen PN. Optimizing multiplex and LA-PCR with betaine. *Trends in Biomedical Sciences* 1997; **22**: 225-226.

Henke W, Herdel K, Jung K, Schnorr D, Loening S A. Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Research* 1997; **25**: 3957-3958.

Herbig K, Chiang E-P, Rees L-R, Hills J, Shane B, Stover P J. Cytoplasmic serine hydroxmethytransferase mediates competition between folate-dependent deoxyribonucleotide and S-adensoylmethionine biosynthesis. *Journal of Biological Chemistry* 2002; **41**: 38381-38389.

Hijiya N, Hudson M H, Lensing S, Zacher M, Onciu M, Behm F G, Razzouk B I, Ribeiro R C, Rubnitz J E, Sandlund J T, Rivera G K, Evans W E, Relling M V, Pui C-H. Cumulative Incidence of Secondary Neoplasms as a First Event After Childhood Acute Lymphoblastic Leukemia. *Journal of the American Medical Association* 2007; **297**:1207-1215.

Hiraoka M, Kato K, Saito Y, Yasuda K, Kagawa Y. Gene-nutrient and gene-gene interactions of controlled folate intake by Japanese women. *Biochemical and Biophysical Research Communication* 2004; **316**: 1210-1216.

Hoelzer D, Gökbuget N, Ottmann O, Pui C-H, Relling M V, Applebaum F R, van Dongen J J M, Szczepański T. Acute lymphoblastic leukamia. *Haematology* 2002; (Am Soc Hematol Educ Program): 162-92.

Hoffbrand A V, Tripp E, Catovsky D, Das K C. Transport of methotrexate into normal haemopoietic cells and into leukaemic cells and its effects on DNA synthesis. *British Journal of Haematology* 1973; **25**: 497-511.

Hoffmann K, Firth M J, Beesley A H, Freitas J R, Ford J, Senanayake S, de Klerk N H, Baker D L, Kees U R. Prediction of relapse in paediatric pre-B acute lymphoblastic leukaemia using a three-gene risk index. *British Journal of Haematology* 2008; **140**: 656-664.

Hogarth L A, Hall A G. Increased BAX expression is associated with an increased risk of relapse in childhood acute lymphoblastic leukemia. *Blood* 1999; **93**: 2671-2678.

Hooijberg J H, Broxterman H J, Kool M, Assaraf Y G, Peteres G J, Noordhuis P, Scheper R J, Borst P, Pinedo H M, Jansen G. Antifolate resistance mediated by the multidrug resistance proteins MRP1 and MRP2. *Cancer Research* 1999; **59**: 2532-2535.

Hooijberg J H, Peters G J, Assaraf Y G, Kathmann I, Priest D G, Bunni M A, Veerman A J P, Scheffer G L, Kaspers G J L, Jansen G. The role of multidrug resistance proteins MRP1, MRP2 and MRP3 in cellular folate homeostasis. *Biochemical Pharmacology* 2003; **65**: 765-771.

Horie N, Takeishi K. Identification of functional elements in the promoter region of the human gene for thymidylate synthase and nuclear factors that regulate the expression of the gene. *Journal of Biological Chemistry* 1997; **272**: 18375-18381.

Hornung N, Stengaard-Pedersen K, Ehrnrooth E, Ellingsen T, Poulsen J H. The effects of low-dose methotrexate on thymidylate synthetase activity in human peripheral blood mononuclear cells. *Clinical and Experimental Rheumatology* 2000; **18**: 691-698.

Horton H R, Moran L A, Ochs R S, Rawn J D, Scrimgeour K G. *Principles of Biochemistry* (3rd Edition). Neil Patterson Publishers, Prentice-Hall: Englewood Cliffs, NJ. 1993; 582.

Horton H R, Moran L A, Scrimgeour K G, Perry M D, Rawn J D. *Principles of Biochemistry* (4th Edition) Pearson Education International: New Jersey. 2006: 208-209.

Hu N, Wang C, Hu Y, Yang H H, Kong L-H, Lu N, Su H, Wang Q-H, Goldstein A, Buetow K H, Emmert-Buck M R, Taylor P R, Lee M P. Genome-wide loss of heterozygosity and copy number alteration in esophageal squamous cell carcinoma using Affymetrix GeneChip Mapping 10K array. *BioMed Central Genomics* 2006; **7**: 299. Available from *http:www.biomedcentral.com/1471-2164/7/299.*

Huang S, Delcamp T J, Tan X, Smith P L, Predergast N J, Freisheim J H. Effects of conversion of an invariant tryptophan residue to phenylalanine on the function of human dihydrofolate reductase. *Biochemistry* 1989; **28**: 471-478.

Huang S, Appleman J R, Tan X, Thompson P D, Blakely R L, Sheridan R P, Venkataraghavan R, Freisheim J H. Role of lysine-54 in determining cofactor specificity and binding in human dihydrofolate reductase. *Biochemistry* 1990; **29**: 8063-8069.

Huber C G, Berti G N. Detection of partial denaturation in AT-rich DNA fragments by ion-pair reversed-phase chromatography. *Analytical Chemistry* 1996; **68**: 2959-2965.

Huber C G, Oefner P J, Bonn G K. High-performance liquid chromatography separation of detritylated oligonucleotides on highly cross-linked poly-(styrene-divylbenzene) particles. *Journal of Chromatography* 1992; **599**: 113-118.

Huber C G, Oefner P J, Bonn G K. High-performance liquid chromatography of oligonucleotides on nonporous alkylated styrne-divinylbenzene copolymers. *Analytical Biochemistry*1993a; **212**: 351-358.

Huber C G, Oefner P J, Preuss E, Bonn G K. High-resolution liquid chromatography of DNA fragments on non-porous poly(styrene-divinylbenzene) particles. *Nucleic Acids Research* 1993b; **21**: 1061-1066.

Huber C G, Oefner P J, Bonn G K. Rapid analysis of biopolymers on modofoed non-porous polystyrene-divinylbenzene particles. *Chromatographia* 1993c; **37**: 653-658.

Huber C G, Oefner P J, Bonn G K. Rapid and accurate sizing of DNA fragments by ion-pairing chromatography on alkylated nonporous poly(styrene-divinylbenzene) particles. *Analytical Chemistry* 1995; **67**: 578-585.

Huber C G, Premstaller A, Xiao W, Oberacher H, Bonn G K, Oefner P J. Mutation detection by capillary denaturing high-performance liquid chromatography using monolithic columns. *Journal of Biochemical and Biophysical Methods* 2001; **47**: 5-19.

Iacopetta B, Grieu F, Joseph D, Elsaleh H. A polymorphism in the enhancer region of the thymidylate synthase promoter influencing the survival of colorectal cancer patients treated with 5-fluorouracil. *British Journal of Cancer* 2001; **85**: 827-830.

Invernizzi R, Comincini S, Travaglino E, Facoetti A, Ramajoli I, Nano R. Defective expression of the dihydrofolate reductase gene in patients with 5q-syndrome. *Haematologica* 2003; **88**: 471-474.

Iqbal M P, Burney I A, Sultana F, Mehboobali N, Siddiqui T. Increased levels of multiple forms of dihydrofolate reductase in peripheral blood leucocytes of cancer patients receiving haematopoietic colony-stimulating factors: interim analysis. *Experimental and Molecular Medicine* 2000; **32**: 84-87.

Irving J A E, Bloodworth L, Brown N P, Case M C, Hogarth L A, Hall A G. Loss of heterozygosity in childhood acute lymphoblastic leukemia detected by genome-wide microarray single nucleotide polymorphism analysis. *Cancer Research* 2005; **65**: 3053-3058.

Ito C, Kumagai M-A, Manabe A, Cousten-Smith E, Raimondi S C, Behm F G, Murti K G, Rubnitz J E, Pui C-H, Campana D. Hyperdiploid acute lymphoblastic leukaemia with 51 to 65 chromosomes: A distinct biological entity with a marked propensity to undergo apoptosis. *Blood* 1999; **93**: 315-320.

Jacky P B, Beek B, Sutherland G R. Fragile sites in chromosomes: possible model for the study of spontaneous chromosome breakage. *Science* 1983; **220**: 69-70.

Jansen G, Westerhoff G R, Jarmuszewski M J A, Kathmann I, Rijksen G, Schornagel J H. Methotrexate transport in variant human CCRF-CEM leukemia cells with elevated levels of the reduced folate carrier. *Journal of Biological Chemistry* 1990; **265**: 18272-18277.

Jansen G, Mauritz R, Drori S, Sprechert H, Kathmann I, Bunni M, Priest D G, Noordhuist P, Schornagel J H, Pinedo H M, Peters G J, Assaraf Y G. A structurally altered human reduced folate carrier with increased folic acid transport mediates a novel mechanism of antifolate resistance. *Journal of Biological Chemistry* 1998; **273**: 30189-30198.

Jansen G. Receptor- and carrier-mediated transport systems for folates and antifolates; In; **Antifolate drugs: Basic Research and Clinical Practice**. Jackman A L (Editor). Humana Press Inc, Totowa, NJ. 1999; 293-321.

Jenkins F L, Turner S L. Use of deoxyinosine in PCR to improve amplification of GC-rich DNA. *BioTechniques* 1995; **19**: 48-52.

Johnson W G, Stenroos E S, Spychala J R, Chatkupt S, Ming S X. New 19 bp deletion polymorphism in intron-1 of dihydrofolate reducatase (DHFR): A risk factor for spina bifida acting in mothers during pregnancy? *American Journal of Medical Genetics* 2004; **124A**: 339-345.

Johnson W G, Scholl T O, Spychala J R, Buyske S, Stenroos E S, Chen X. Common dihydrofolate reductase 19-base pair deletion allele: a novel risk factor for preterm delivery. *American Journal of Clinical Nutrition* 2005; **81**: 664-668.

Jolivet J, Cowan K H, Curt G A, Clendeninn N J, Chabner B A. The pharmacology and clinical use of methotrexate. *New England Journal of Medicine* 1983; **309**: 1094-1104.

Jolkowska J, Derwich K, Dawidowska M. Methods of minimal residual disease (MRD) detection in childhood haematological malignancies. *Journal of Applied Genetics* 2007; **48**: 77-83.

Jones A C, Austin J, Hansen, Hoogendoorn B, Oefner P J, Cheadle J P, O'Donovan M C. Optimal temperature selection for mutation detection by denaturing HPLC and comparisons of single-stranded conformation polymorphism and heteroduplex analysis. *Clinical Chemistry* 1999; **45**: 1133-1140.

Jost W, Unger K, Schill G. Reverse-phase ion-pair chromatography of polyvalent ions using olignucleotides as model substances. *Analytical Biochemistry* 1982; **119**: 214-223.

Kadan-Lottick N S, Ness K K, Bhatia S, Gurney J G. Survival variability by race and ethnicity in childhood acute lymphoblastic leukaemia. *Journal of the American Medical Association* 2003; **290**: 2008-2014.

Kager L, Cheok M, Yang W, Zaza G, Cheng Q, Panetta J C, Pui C-H, Downing J R, Relling M V, Evans W E. Folate pathway gene expression differs in subtypes of acute lymphoblastic leukaemia and influences methotrexate pharmacodynamics. *Journal of Clinical Investigation* 2005; **115**: 110-117.

Kaneda S, Takeishi K, Ayusawa D, Shimizu K, Seno T, Altman S. Role in translation of a triple tandemly repeated sequence in the 5'-untranslated region of human thymidylate synthase mRNA. *Nucleic Acids Research* 1987; **15:** 1259-1269.

Kastrup I B, Worm J, Ralfkiaer E, Hokland P, Guldberg P, Gronbaek K. Genetic and epigenetic alterations of the reduced foalte carrier in untreated diffuse large B-cell lymphoma. *European Journal of Haematology* 2007; **80**: 61-66.

Kaufman Y, Drori S, Cole P D, Kamen B A, Sirota J, Ifergan I, Arush M W B, Elhasid R, Sahar D, Kaspers G J L, Jansen G, Matherly L H, Rechavi G, Toren A, Assaraf Y G. Reduced folate carrier mutations are not the mechanism underlying methotrexate resistance in childhood acute lymphoblastic leukaemia. *Cancer* 2003 (ePub ahead of printed copy in 2004); **100**: 773-782.

Kaufman Y, Ifergan I, Rothem L, Jansen G, Assaraf Y G. Coexistence of multiple mechanisms of PT523 resistance in human leukemia cells harbouring 3 reduced folate carrier alleles: transcriptional silencing, inactivating mutations, and allele loss. *Blood* 2006; **107**: 3288-3294.

Kawakami K, Watanabe G. Identification and functional analysis of single nucleotide polymorphism in the tandem repeat sequence of thymidylate synthase gene. *Cancer Research* 2003; **63**: 6004-6007.

Ke S-H and Wartell R M. Influence of nearest neighbor sequence on the stability of base pair mismatches in long DNA: determination by temperature-gradient gel electrophoresis. *Nucleic Acids Research* 1993: **21**; 5137-5143.

Kim J-S, Shane B. Role of folylpolyglutamate synthetase in the metabolism and cytotoxicity of 5-deazaacyclotetrahydrofolate, an anti-purine drug. *Journal of Biological Chemistry* 1994; **269**: 9714-9720.

Kim Y-I. Folate and carcinogenesis: evidence, mechanisms, and implications. *Journal of Nutritional Biochemistry* 1999; **10**: 66-88.

Kishi S, Griener J, Cheng C, Das S, cook E H, Pei D, Hudson M, Rubnitz J, Sandlund J T, Pui C-H, Relling M V. Homocysteine, pharmacogenetics, and neurotoxicity in children with leukaemia. *Journal of Clinical Oncology* 2003; **21**: 3084-3091.

Khokhar N Z, Lam A F Y, Rusch V W, Sirotnak F M. Despite some expression of folate receptor α in human mesothelioma cells, Internalisation of methotrexate is predominately carrier mediated. *General Thoracic Surgery* 2002; **123**: 862-868.

Kohlmeier M, da Costa K-A, Fisher L M, Zeisel S H. Genetic variation of folate-mediated one-carbon transfer pathway predicts susceptibility to choline deficiency in humans. *Proceedings of the National Academy of Science* 2005; **102**: 16025-16030.

Kosaki K, Yoshihashi H, Ohashi Y, Kosaki R, Suzuki T, Matsuo N. Fluorescence-based DHPLC for allelic quantification by single-nucleotide primer extension. *Journal of Biochemical and Biophysical Methods* 2001; **47**: 111-119.

Kosaki K, Udaka T, Okuyama T. DHPLC in clinical molecular diagnostic services. *Molecular Genetics and Metabolism* 2005; **86**: 117-123.

Krajinovic M, Costea I, Chiasson S. Polymorphisms of the gene and outcome of acute lymphoblastic leukaemia. *Lancet* 2002; **359**: 1033-1034.

Krause J. Stem cells. *Whyfiles.org/127stem_cell/3.html* Visited 28/1/2006.

Kraus R J, Shadley L, Mertz J E. Nuclear factor 1 family members mediate repression of the BK virus late promoter. *Virology* 2001; **287**: 89-104.

Kruh G D, Zeng H, Rea P A, Liu G, Chen Z-S, Lee K, Belinsky M G. MRP subfamily transporters and resistance to anticancer agents. *Journal of Bioenergetics and Biomembranes* 2001; **33**; 493-501.

Kuklin A I, Haefele R, Taylor P D, Gjerde D T, Hecker K, Munson K. Mutation detection using DNA chromatography (DHPLC). *American Journal of Human Genetics* 1999; **65 (supple 5)**: 1206.

Kuchinskaya E, Heyman M, Nordgren A, Schoumans J, Staaf J, Borg A, Soderhall S, Grander D, Nordenskjold M, Blennow E. Array-CGH reveals hidden gene dose changes inchildren with acute lymphoblastic leukaemia and a normal or failed karyotype by G-banding. *British Journal of Haemology* 2008; **140**: 572-577.

Kwok P-Y, Chen X. Detection of single nucleotide polymorphisms. *Current Issues in Molecular Biology* 2003; **5**: 43-60.

Lapenta V, Sossi V, Gosset P, Vaysssettes C, Vitali T, Rabatel N, Tassone F, Blouin J-L, Scott H S, Antonarakis S E, Crau N, Brahe C. Construction of a 2.5-Mb integrated physical and gene map of distal 21q22.3. *Genomics* 1998; **49**: 1-13.

Lathouder S de, Gerards A H, de Groot E R, Valkhof M G, Dijkmans B A, Aarden L A. Bioassay for detection of methotrexate in serum. *Scandinavian Journal of Rheumatology* 2004; **33**: 167-173.

LePecq J-B, Paoletti C. A fluorescent complex between ethidium bromide and nucleic acids. *Journal of Molecular Biology* 1967; **27**: 87-106.

Leukaemia Research Fund. Statistics. *http://www.lrf.org.uk.* Visited 17/4/2008.

Leclerc D, Wilson A, Dumas R, Gafuik C, Song D, Watkins D, Heng H H Q, Rommens J M, Scherer S W, Rosenblatt D S, Gravel R A. Cloning and mapping of a cDNA for methionine synthase reductase, a flavoprotein defective in patients with homocystinuria. *Proceedings of the National Academy of Science* 1998; **95**: 3059-3064.

Levy A S, Sather H N, Steinherz P G, Sowers R, La M, Moscow J A, Gaynon P S, Uckun F M, Bertino J R, Gorlick R. Reduced folate carrier and dihydrofolate reductase expression in acute lymphocytic leukemia may predict outcome: a Children's Cancer Group study. *Journal of Pediatric Hematology/Oncology* 2003; **25**: 688-695.

Lewis W S, Cody V, Galitsky N, Lufts J R, Pangborn W, Chunduru S K, Spencer H T, Appleman J R, Blakley R L. Methotrexate-resistant variants of human dihydrofolate reductase with substitutions of leucine 22. *Journal of Biological Chemistry* 1995; **270**: 5057-5064.

Liang H, Samanta S, Nagarajan L. SSBP2, a candidate tumor suppressor gene, induces growth arrest and differentiation of myeloid leukemia cells. *Oncogene* 2005; **24**:2625-2634.

Liani E, Rothem L, Bunni M A, Smith C A, Jansen G, Assaraf Y G. Loss of folylpoly-γ-glutamate synthetase activity is a dominant mechanism of resistance to polyglutamylation-dependent novel antifolates in multiple human leukaemia sublines. *International Journal of Cancer* 2003; **103**: 587-599.

Liu M R, Pan K F, Wang Y, Lu Y Y. Influence of DNA polymerase on mutation screening of denaturing high-performance liquid chromatography. *Ai Zheng* 2002; **21**: 1160-1163. Article in Chinese, abstract in English.
Liu W, Smith D I, Rechtzigel K J, Thibodeau S N, James C D. Denaturing high performance liquid chromatography (DHPLC) used in the detection of germline and somatic mutations. *Nucleic Acids Research* 1998; **26**: 1396-1400.

Liu X Y, Matherly L H. Functional interaction between arginine-133 and aspartate-88 in the human reduced folate carrier: evidence for a charge-pair association. *Biochemical Journal* 2001; **358**: 511-516.

Liu X Y, Matherly L H. Analysis of membrane topology of the human reducee folate carrier protein by hemagglutinin epitope insertion and scanning glycosylation insertion mutagenesis. *Biochimica et Biophysica Acta* 2002; **1564**: 333-342.

Liu X Y, Witt T L, Matherly L H. Restoration of high level transport activity by human reduced folate carrier/ThTr1 chimeric transporters: Role of the transmembrane domain 6/7 linker region in reduced folate carrier protein. *Biochemical Journal* 2003; **369**: 31-37.

Ljungman M, Hanawalt P C. Localized torsional tension in the DNA of human cells. *Proceedings of the National Academy of Science* 1992; **89**: 6055-6059.

Lodish H, Baltimore D, Berk A, Zipursky S L, Matsudaira P, Darnell J. Sorting of plasma membrane, secretory and lysosomal proteins. **Molecular Cell Biology (3rd Edition).** Scientific American Books 1995; 697-699.

Lohrum M A E, Vousden K H. Regulation and function of the p53-related proteins: same family different rules. *Trends in Cell Biology* 2000; **10**: 197-202.

Loo R W, Clarke D M. Prolonged association of temperature-sensitive mutants of human P-glycoprotein with calnexin during biogenesis. *Journal of Biological Chemistry* 1994; **269**: 28683-28689.

Lu Y, Kham S K-Y, Foo T-C, Hany A, Quah T-C, Yeoh A E-J. Genotyping of eight polymorphic genes encoding drug-metabolizing enzymes and transporters using a customized oligonucleotide array. *Analytical Biochemistry* 2007; **360**: 105-113.

McCloskey D E, McGuire J J, Russell C A, Rowan B G, Bertino J R, Pizzorno G, Mini E. Decreased folylpolyglutamate synthetase activity as a mechanism of methotrexate resistance in CCRF-CEM human leukemia sublines. *Journal of Biological Chemistry*; **266**: 6181-6187.

McHale C M, Wiemels J L, Zhang L, MA X, Buffler P A, Guo W, Loh M L, Smith M T. Prenatal origin of *TEL-AML1*- positive acute lymphoblastic leukemia in children born in California. *Gene, Chromosomes and Cancer* 2003; **37**: 36-43.

McKenzie S B. **Textbook of Hematology (2nd Edition)**. Williams & Wilkins: Baltimore. 1996.

McKusick V A. Solute carrier family 19, member 1: SLC19A1. *Online Mendelian Inheritance in Man. http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=600424* John Hopkins University. Visited 3/12/2005.

MacGregor J T, Wehr C M, Hiatt R A, Peters B, Tucker J D, Langlois R G, Jacob R A, Jensen R H, Yager J W, Shigenaga M K, Frei B, Eynon B P, Ames B N. 'Spontaneous' genetic damage in man: evaluation of interindividual variability, relationship among markers of damage, and influence of nutritional status. *Mutation Research* 1997; **377**: 125-135.

Mandola M V, Stoehlmacher J, Muller-Weeks S, Cesarone G, Yu M C, Lenz H-J, Ladner R D. A novel single nucleotide polymorphism within the 5' tandem repeat polymorphism of the *thymidylate synthase* gene abolishes USF-1 biding and alters transcriptional activity. *Cancer Research* 2003; **63**: 2898-2904.

Marchant J S, Subramanian V S, Parker I, Said H. Intracellular trafficking and membrane targeting mechanisms of the human reduced folate carrier in mammalian epithelial cells. *Journal of Biological Chemistry* 2002; **277**: 33325-33333.

Marinaki A M, Arenas M, Duley J A, Shobowale-Bakre E P, Ansari A, Sanderson J D, Fairbanks L D. **Programme & Abstracts, Joint 11th International and 9th European Symposium on Purines and Pyrimidines in man:** 2003; PP03: P33a
.

Marmur J, Lane D. Strand separation and specific recombination in deoxyribonucleic acids: biological studies. *Proceedings of the National Academy of Science* 1960; **46**: 453-461.

Maslak P. Pediatric Pre B ALL.
http://www.ashimagebank.org/cgi/collection/b-cell_all Visited 28/1/06.

Maslak P. Precursor T-cell ALL. *http://www.ashimagebank.org/cgi/content/full/2004/0223/101018/F1* Visited 28/1/06.

Masson E, Relling M V, Synold T W, Liu W, Schuetz J D, Sandlund J T, Pui C H, Evans W E. Accumulation of methotrexate polyglutamates in lymphoblasts is a determinant of antileukemia effects in vivo: a rationale for high-dose methotrexate. *Journal of Clinical Investigation* 1996; **97**: 73-80.

Matherly L H. Molecular and cellular biology of the human reduced folate carrier. *Progress in Nucleic Acid Research and Molecular Biology* 2001; **67**: 131-162.

Matherly L H, Angeles S H. Role of N-glycosylation in the structure and function of the methotrexate membrane transporter from CCRF-CEM human lymphoblastic cells. *Biochemical Pharmacology* 1994; **47**: 1094-1098.

Matherly L H, Goldman I D. Membrane transport of folates. *Vitamins and Hormones* 2003; **66**: 403-456.

Matherly L H, Taub J W. Molecular and cellular correlates of methotrexate response in childhood acute lymphoblastic leukemia. *Leukemia and Lymphoma* 1999; **35**: 1-20.

Matherly L H, Taub J W, Ravindranath Y, Proefke S A, Wong S C, Gimotty P, Buck S, Wright J E, Rosowsky A. Elevated dihydrofolate reductase and impaired methotrexate transport as elements in methotrexate resistance in childhood acute lymphoblastic leukemia. *Blood* 1995; **85**: 500-509.

Matherly L H, Taub J W, Wong S C, Simpson P M, Ekizian R, Buck S, Williamson M, Amylon M, Pullen J, Camitta B, Ravindranath Y. Increased frequency of expression of elevated dihydrofolate reductase in T-cell versus B-precursor acute lymphoblastic leukemia in children. *Blood* 1997; **90**: 578-589.

Matheson E C, Hogarth L A, Case M C, Irving J A E, Hall A G. DHFR and MSH3 co-amplification in childhood acute lymphoblastic leukaemia, *in vitro* and *in vivo*. *Carcinogenesis* 2007; **28**: 1341-1346.

Maurer B J, Barker P E, Masters J N, Ruddle F H, Attardi G. Human dihydrofolate reductase gene is located in chromosome 5 and is unlinked to the related pseudogenes. *Proceedings of the National Academy of Science* 1984; **81**: 1484-1488.

Mauritz R, Peters G J, Priest D G, Assaraf Y G, Drori S, Kathmann I, Noordhuis P, Bunni M A, Rosowsky A, Schornagel J H, Pinedo H M, Jansen G. Multiple mechanisms of resistance to methotrexate and novel antifolates in human CCRF-CEM leukemia cells and their implications for folate homeostasis. *Biochemical Pharmacology* 2002; **63**: 105-115.

Mauritz R, Peters G, Kathmann I, Teshale H, Noordhuis P, Comijn E M, Pinedo H M, Jensen G. Dynamics of antifolate transport via the reduced folate carrier and the membrane folate receptor in murine leukaemia cells in vitro and in vivo. *Cancer Chemotherapy and Pharmacology* 2008; E-pub ahead of print DOI 10.1007/s00280-008-0683-0.

Meaburn E, Butcher L M, Schalkwyk L C, Plomin R. Genotyping pooled DNA using 100k SNP microarrays: a step towards genomewide association scans. *Nucleic Acids Research* 2006; **34**: e28.

Melchior (Jr) W B and Von Hippel P. Alteration of the relative stability of dA•dT and dG•dC base pairs in DNA. *Proceedings of the National Academy of Science* 1973: **70**; 298-302.

Merola P R, Sowers R, Yang R *et al.* Reduced folate carrier sequence alterations are not common in leukaemia/lymphoma samples. *Proceedings of the American Association of Cancer Research* 2002; **43**: 60-61.

Mesner L D, Hamlin J L. Specific signals at the 3'-end of the DHFR gene define one boundry of the downstream origin of replication. *Genes and Development* 2005; **19**: 1053-1066.

Meyer K, Fredrikson A, Ueland P M. High-level multiplex genotyping of polymorphisms involved in folate or homocysteine metabolism by matrix-assisted laser desorpttion/ionisation mass spectrometry. *Clinical Chemistry* 2004; **50**: 391-402.

Mi S, Lu J, sun M, Li Z, Zhang H, Neilly M B, Wang Y, Qian Z, Jin J, Zhang Y, Bohlander S K, LeBeau M M, Larson R A, Golub R T, Rowley J D, Chen J. MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia. *Proceedings of the National Academy of Science* 2007; **104**: 19971-19976.

Miller D V, Leontovich A A, Lingle W L, Suman V J, Mertens L L, Lillie J, Ingalls K A, Perez E A, Ingle J N, Couch F J, Visscher D W. Utilizing Nottingham Prognostic Index in microarray gene expression profiling of breast carcinomas. *Modern Pathology* 2004; **17**: 756-764.

Mirinaki A M, Arenas M, Duley J A, Shobowale-Bakere L T, Ansari A, Sanderson J D, Fairbanks L D. Polymorphism in the MTHFR gene effects thiopurine methyltransferase activity. **Programme & Abstracts, Joint 11th International and 9th European Symposium on Purines and Pyrimidines in Man.** 2003: PP03

Mittal R, Mottl H, Memec J. Acute transient cerebral toxicity associated with administration of high-dose methotrexate. *Medical Principles and Practice* 2005; **14**: 202-204.

Miyachi H, Takemura Y, Kobayashi H, Ando K, Ando Y. Differential alterations of dihydrofolate reductase gene in human leukemia cell lines made resistant to various folate analogues. *Japanese Journal of Cancer Research* 1993; **84**: 9-12.

Moorman A V, Richards S M, Marineau M, Cheung K L, Robinson H M, Jalali G R, Broadfield Z J, Harris R L, Taylor K E, Gibson B E S, Hann I M, Hill F G H, Kinsey S E, Eden T O B, Mitchell C D, Harrison C J. Outcome heterogeneity in childhood high-hyperdiploid acute lymphoblastic leukaemia. *Blood* 2003; **102**: 2756-2762.

Moorman A V, Richards S M, Robinson H M, Stefford J C, Gibson B E S, Kinsey S E, Eden T O B, Vora A J, Mitchel C D, Harrison C J. Prognosis of children with acute lymphoblastic leukemia (ALL) and intrachromosomal amplification of chromomsome 21 (iAMP21). *Blood* 2008; **109**: 2327-2330.

Mori H, Colman S M, Xiao Z, Ford A M, Healy L E, Donaldson C, Hows J M, Navarrete C, Greaves M. Chromosomes translocations and covert leukemic clones are generated during normal fetal development. *Proceedings of the National Academy of Science* 2002; **99**: 8242-8247.

Morin I, Devlin A M, Leclerc D, Sabbaghian N, Halsted C H, Finnell R, Rozen R. Evaluation of genetic variants in the reduced folate carrier and in glutamate carboxypeptidase II for spina bifida risk. *Molecular Genetics and Metabolism* 2003; **79**: 197-200.

Morrow M, Horton S, Kioussis D, Brady H J M, Williams O. TEL-AML1 promotes development of specific hematopoietic lineages consistent with preleukemic activity. *Blood* 2004; **103**: 3890-3896.

Moscow J A, Gong M, He R, Sgagias M K, Dixon K H, Anzick S L, Meltzer P S, Cowan K H. Isolation of a gene encoding a human reduced folate carrier (RFC1) and analysis of its expression in transport-deficient, methotrexate-resistant human breast cancer cells. *Cancer Research* 1995; **55**: 3790-3794.

Mosquera-Cro M, Helman P, Veroff R, Shuster J, Martin S, Davidson G, Poter J, Harvey R, Hromas R, Andries E, Atlas S, Wilson C, Ar K, Yuexian X, Chen I-M, Carroll A, Camitta B, Willman C. Identification, validation, and cloning of a novel gene (OPAL1) and associated genes highly predictive of outcome in pediatric acute lymphoblastic leukaemia using gene expression profiling. *Blood* 2003; **102**: abstract #1.

MRC UKALL 2003. UK National acute lymphoblastic leukaemia (ALL) trial. Medical Research Council Working Party on Leukaemia in Children. *Version 3,* December 2003. Available at *www.ctsu.ox.ac.uk/projects/ukall2003* Visited 31/3/2006.

Mullighan C G, Goorha S, RAdtke I, Miller C B, Coustan-Smith E, Dalton J D, Girtman K, Mathew S, Ma J, Pounds S B, Su X, Pui C-H, RElling M V, Evans W E, Shurtleff S A, Downing J R. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 2007; **446**: 758-764.

Mullis K B, Faloona F A. Specific synthesis of DNA *in vitro* via a polymerase chain reaction. *Methods in Enzymology* 1987; **155**: 335-350.

Mutter G L Boynton K A. PCR bias in amplification of androgen receptor alleles, a trinucleotide repeat marker used in clonality studies. *Nucleic Acids Research* 1995; **23**: 1411-1418.

Mytelka D S, Chamberlin M J. Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus. *Nucleic Acids Research* 1996; **24**: 2774-2781.

National Cancer Institute. Childhood Acute Lymphoblastic Leukemia (PDQ®):Treatment. *http://www.cancer.gov/cancertopics/pdq/treatment/childALL/healthprofessional.* Visited 29/10/05.

National Human Genome Research Institute. International Consortium Completes Human Genome Project. *http://www.genome.gov/11006929* Visited 22/7/2006.

Nehls S, Snapp E L,Cole N B, Zaal K J M, Kenworthy A K, Roberts T H, Ellenberg J, Presley J F, Siggia E, Lippincott-Schwartz J. Dynamics and retention of misfolded proteins in native ER membranes. *Nature Cell Biology* 2000: **2**: 288-295.

Ng P C, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Research* 2001; **11**: 863-874.

NHGRI. DNA microarray technology. National Human Genome Research Institute. http://www.genome.gov/10000533. Visited 26/5/07.

Nijhout H F, Reed M C, Budu P, Ulrich C M. A mathematical model of the folate cycle. *Journal of Biological Chemistry* 2004; **279**: 55008-55016.

O'Donovan M C, McGuffin P. Evaluation of DHPLC as a tool for mutation detection. *American Journal of Medical Genetics* 1997; **74**: 600-601.

O'Donovan M C, Oefner P J, Roberts S C, Austin J, Hoogendoorn B, Guy C, Speight G, Upadhyaya M, Sommer S S and McGuffin P. Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics* 1998: **52**; 44-49.

Oefner P J. Allelic discrimination by denaturing high-performance liquid chromatography. *Journal of Chromatography B* 2000: **739**; 345-355.

Oefner P J, Bonn G K. High-resolution liquid chromatography of nucleic acids. *American Laboratory* 1994a; **26:** 28C-28J.

Oefner P J, Huber C G, Umlauft F, Berti G-N, Stimpl E, Bonn G K. High-resolution liquid chromatography of fluorescent dye-labeled nucleic acids. *Analytical Biochemistry* 1994b; **223**: 39-46.

Oefner P J, Underhill P A. Comparative DNA sequencing by denaturing high-performance liquid chromatography (DHPLC). *American Journal of Human Genetics* 1995; **suppl 57**: A266.

Oefner P J, Underhill P A. DNA mutation detection using denaturing high-performance liquid chromatography (DHPLC). *Current Protocols in Human Genetics* 1998; **suppl 19**: 7.10.1-7.10.12.

Ohnuma T, Lo R J, Scanlon K J, Kamen B A, Ohnoshi T, Wolman S R, Holland J F. Evolution of methotrexate resistance of human acute lymphoblastic leukemia cell in vitro. *Cancer Research* 1985; **45**: 1815-1822.

O'Neil M J F, McKusick V A. Dihydrofolate reductase; DHFR. *Online Mendelian Inheritance in Man.* http://www.ncbi.nlm.nih.gov/entrez/dispomim.cg?cmd=entrybid=126060. John Hopkins University. Visited 7/02/2006.

Ornstein R L, Rein R, Breen D L and MacElroy R D. *Biopolymers* (1978): **17**; 2356.

Packer B R, Yeager M, Burdett L, Welch R, Beerman M, Qi L, Sicotte H, Staats B, Acharya M, Crenshaw A, Eckert A, Puri V, Gerhard D, Chanock S J. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Research* 2006; **34 Database issue**: D617-D621.

Pakakasama S, Kanchanakamhaeng K, Kajanachumpol S, Udomsubpayakul U, Sirachainan N, Thithapandha A, Hongeng S. Genetic polymorphisms of folate metabolic enzymes and toxicities of high dose methotrexate in children with acute lymphoblastic leukemia. *Annals of Heamatology* 2007; **86**: 609-611.

Pao S S, Paulsen I T, Saier Jr M H. Major facilitator superfamily. *Microbiology and Molecular Biology Reviews* 1998; **62**: 1-34.

Papadopoulou E, Metaxa-Mariatou V, Hatzaki A, Hatzis T, Nasioulas G. The implication of using mutagenic primers in combination with *Taq* polymerase having proofreading activity. *Biologicals* 2004; **32**: 84-87.

Park T, Yi S G, Kang S H, Lee S Y, Lee Y S, Simon R. Evaluation of normalization methods for microarray data. *BioMed Central Bioinformatics* 2003; **4**: 33.

Parle-McDermott A, Pangilinan E, Mills J L, Kirke P N, Gibney E R, Troendle J, O'Leary J L, Molloy A M, Conley M, Scott J M, Brody L C. The 19-bp deletion polymorphism in intron-1 of dihydrofolate reductase (DHFR) may decrease rather than increase risk for spina bifida in the Irish population. *American Journal of Medical Genetics (Part A)* 2007; **143A**: 1174-1180.

Pauletti G, Lai E, Attardi G. Early appearance and long-term persistence of the submicroscopic extrachromosomal elements (amplisomes) containing the amplified DHFR genes in human cell lines. *Proceedings of the National Academy of Science* 1990; **87**: 2955-2959.

Payton S G, Liu M, Ge Y, Matherly L H. Transcriptional regulation of the human reduced folate carrier A1/A2 promoter: Identification of critical roles for the USF and GATA families of transcription factors. *Biochimica et Biophysica Acta* 2005: **1731**: 115-124.

Payton S G, Haska C L, Flatley R M, Ge Y, Matherly L H. Effect sof 5' untranslated region diversity on the posttranscriptional regulation of the human reduced folate carrier. *Biochimica et Biophysica Acta* 2007; **1769**: 131-138.

Pei L J, Ren A G, Zhu H P, Hao L, Zhao W R, Li Z, Hou G W, Zhang B L, Jiang Y Y, Wu L M, Pan Y J, Zhang M L. Study on reduced folate carrier gene (RFC1) polymorphisms in the southern and northern Chinese population. *Zhanghua Liu Xing Bing Xue Za Zhi* 2004; **25**: 499-502. [Abstract] Article in Chinese.

Peters G J, Jansen G. Folate homeostasis and antiproliferative activity of folates and antifolates. *Nutrition* 2001; **17**: 737-738.

Pfeiffer R M, Bura E, Smith A, Rutter J L. Two approaches to mutation detection based on functional data. *Statistics in Medicine*2002; **21**: 3447-3464.

Poplack D G, Acute lymphoblastic leukaemia. In; Pizza P A, Poplack D G (Eds). **Principles and practice of pediatric oncology**. Philadelphia, Lippincott. 1993.

Prendergst N J, Delcamp T J, Smith P L, Freisheim J H. Expression and site-directed mutagenesis of human dihydrofolate reductase. *Biochemistry* 1988; **27**: 3664-3671.

Prendergst N J, Appleman J R, Delcamp T J, Blakley R L, Freisheim J H. Effects of conversion of phenylalanine-31 tp leucine on the function of human dihydrofolate reductase. *Biochemistry* 1989; **28**: 4645-4650.

Pui C-H, Raimondi S C, Hancock M L, Rivera G K, Riberio R C, Mahmoud H H, Sandlund J T, Crist W M, Behm F G. Immunological, cytogenetic, and clinical characterization of childhood acute lymphoblastic leukaemia with the t(1;19)(q23;p13) or its derivative. *Journal of Clinical Oncology* 1994; **12**: 2601-2606.

Pui C-H, Evans W E. Acute lymphoblastic leukeami. *New England Journal of Medicine* 1998; **339**: 605-615.

Pui C-H, Boyett J M, Relling M V, Harrison P L, Rivera G K, Behm F G, Sandlund J T, Ribeiro R C, Rubnitz J E, Gajjar A, Evans W E. Sex differences in prognosis for children with acute lymphoblastic leukaemia. *Journal of Clinical Oncology* 1999; **17**: 818-824.

Pui C-H, Gaynon P S, Boyett J M, Chessells J M, Baruchel A, Kamps W, Silverman L B, Biondi A, Harms D O, Vilmer E, Schrappe M, Camitta B. Outcome of treatment in childhood acute lymphoblastic leukaemia with rearrangements of the 11q23 chromosomal region. *Lancet* 2002; **359**: 1909-1915.

Pui C-H, Relling M V, Downing J R. Mechanisms of disease: Acute lymphoblastic leukemia. *New England Journal of Medicine* 2004a: **350**: 1535-1548.

Pui C-H, Sandlund J T, Pei D, Campana D, Rivera G K, Ribeiro R C, Rubnitz J E, Bassem I, Razzouk S C, Scott C H, Hudson M M, Cheng C, Kun L E, Raimondi S C, Behm F G, Downing J R, Relling M V, Evans W E. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIB at St Jude Children's Research Hospital. *Blood* 2004b; **104**: 2690-2696.

Pui C-H. Neonatal leukemia: A nemesis for pediatric oncologists? [Editorial] *Pediatric Blood Cancer* 2005 (e-version DOI 10.1002/pbc.20607 ahead of printed version in 2006): **47**: 234-235.

Pui C-H, Robison L L, Look A T. Acute lymphoblastic leukaemia. *Lancet* 2008; **371**: 1030-1043.

Quinlivan E P, Davis S R, Shelnutt K P, Henderson G N, Ghandour H, Shane B, Selhub J, Bailey L B, Stacpoole P W, Gregory J F. Methylenetetrahydrofolate reductase 677C→T polymorphism and folate status affects one-carbon incorporation into human DNA deoxynucleosides. *Journal of Nutrition* 2005; **135**: 389-396.

Rady P L, Szucs S, Matalon R K. Genetic polymprophism (G80A) or reduced folate carrier gene in ethnic populations [Letter]. *Molecular Genetics and Metabolism* 2001; **73**: 285-286.

Raimondi S C, Roberson P K, Pui C-H, Behm F G, Rivera G K. Hyperdiploid (47-50) acute lymphoblastic leukemia in children. *Blood* 1992; **79**: 3245-3252.

Rajgopal A, Sierra E E, Zhao R, Goldman I D. Expression of the reduced folate carrier SLC19A1 in IEC-6 cells result in two distinct transport activities. *American Journal of Physiology: Cell Physiology* 2001; **281**: C1579-C1586.

Rajagopalan P T, Benkovic S J. Preorganization and protein dynamics in enzyme catalysis. *Chemical Record* 2002; **2**: 24-36.

Rao N A, Talwar R, Savithri H S. Molecular organization, catalytic mechanism and function of serine hydroxmethyltransferase – a potential target for cancer chemotherapy. *International Journal of Biochemistry and Cell Biology* 2000; **32**: 405-416.

Rayl E A, Moroson B A, Beardsley G P. The human *pur*H gene product, 5-aminoimidazole-4-caboxamide ribonucletotide formyltransferase/IMP cyclohydrolase. *Journal of Biological Chemistry* 1996; **271**: 2225-2233.

Rees W A, Yager T D, Korte J, von Hippel P H. Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* 1993; **32**: 137-144.

Relton C L, Pearce M S, Burn J, Parker L. An investigation of folate-related genetic factors in the determination of birthweight. *Paediatric and Perinatal Epidemiology* 2005; **19**: 360-367.

Renwick S B, Snell K, Bauman U. The crystal structure of human cytosolic serine hydroxymethyltransferase: a target for cancer chemotherapy. *Structure* 1998; **6**: 1105-1116.

Reploeg M D, Storch G A, Clifford D B. BK virus: a clinical review. *Clinical Infectious Diseases* 2001; **33**: 191-202.

Robards K, Haddad P R and Jackson P E. **Principles and practices of modern chromatographic methods**. Academic Press: London. 1994; 344-351.

Roche Molecular Systems. AmpliTaq Gold® N808-0240 product insert. Applied Biosystems, 2000.

Rod T H, Brooks C L. How dihydrofolate reducatase facilitates protonation of dihydrofolate. *Journal of the American Chemical Society* 2003; **125**: 8718-8719.

Ross M. Medium sized lymphocyte. Science Photo Library. *http://www.sciencephoto.com/search/searchLogic.html?frontpage=1&searchstring=lymphocyte&country=67* Visited 28/1/2006

Ross M E, Zhou X, Song G, Shurtleff S A, Girtman K, Williams W K, Liu H-C, Mahfouz R, Raimondi S C, Lenny N, Patel A, Downing J R. Classification of pediatric acute lymphoblastic leukaemia by gene expression profiling *Blood* 2003; **102**: 2951-2959.

Rothem L, Ifergan I, Kaufman Y, Priest D G, Jansen G, Assaraf Y G. Resistance to multiple novel antifolates is mediated via defective drug transport resulting from clusetered mutations in the reduced folate carrier gne in human leukaemia cell lines. *Biochemical Journal* 2002; **367**: 741-750.

Rothem L, Aronheim A, Assaraf Y G. Alterations in the expression of transcription factors and the reduced folate carrier as a novel mechanism of antifolate resistance in human leukemia cells. *Journal of Biological Chemistry* 2003; **278**: 8935-8941.

Rothem L, Stark M, Kaufman Y, Mayo L, Assaraf Y G. Reduced folate carrier gene silencing in multiple antifolate-resistant tumour cell lines is due to simultaneous loss of function of multiple transcription factors but not promoter methylation. *Journal of Biological Chemistry* 2004; **279**: 374-384.

Rothenberg S P, Jabal M P. Human leukemia and normal leukocytes contain species of immunoreactive but nonfunctional dihydrofolate reductase. *Proceedings of the National Academy of Science* 1982; **79**: 645-649.

Rots M G, Pieters R, Peters G J, Noordhuis P, van Zantwijk C H, Kaspers G J L, Hahlen K, Creutzig U, Veerman A J P, Jansen G. Role of folylpolyglutamate synthetase and folylpolyglutamate hydrolase in methotrexate accumulation and polyglutamylation in childhood leukemia. *Blood* 1999; **93**: 1677-1683.

Rots M G, Pieters R, Peters G J, Noordhuis P, van Zantwijk CH, Henze G, Janka-Schaub G E, Veerman A J P, Jansen G. Methotrexate resistance in relapsed childhood acute lymphoblastic leukaemia. *British Journal of Haematology* 2000a; **109**: 629-634.

Rots M G, Pieters R, Kaspers G J L, Veerman A J P, Peters G J, Jansen G. Classification of *ex vivo* methotrexate resistance in acute lymphoblastic and myeloid leukaemia. *British Journal of Haematology* 2000b; **110**: 791-800.

Rots M G, Willey J C, Jansen G, van Zantwijk C H, Noordhuis P, DeMuth J P, Kuiper E, Veerman A J P, Pieters R, Peters G J. mRNA expression levels of methotrexate resistance-related proteins in childhood leukemia as determined by a standardized competitive template-based RT-PCR method. *Leukemia* 2000c; **14**: 216-2175.

Roy K, Mitsugi K, Sirlin S, Shane B, Sirotnak F M. Different antifolate-resistant L1210 cell variants with either increased or decreased folylpolyglutamate synthetase geen expression at the level of mRNA transcription. *Journal of Biological Chemistry* 1995; **270**: 26918-26922.

Russell P J. **iGenetics**. Pearson Education Inc, San Francisco, 2002. Chapter 7; 190-192.

Ryder S. 500K microarrays used to detect tiny chromosomal changes in ovarian tumours. *Microarray Bulletin* 2007; **3**: Winter 2007.

Sadlish H, Murray R C, Williams M R, Flintoff W F. Mutations in the reduced-folate carrier affect protein localization and stability. *Biochemical Journal* 2000; **346**: 509-518.

Sadlish H, Williams F M, Flintoff W F. Cytoplasmic domains of the reduced folate carrier are essential for trafficking, but not function. *Biochemical Journal*. 2002a; **364**: 777-786.

Sadlish H, Williams F M R, Flintoff W F. Functional role of arginine 373 in substrate translocation by the reduce folate carrier. *Journal of Biological Chemistry* 2002b; **277**: 42105-42112.

Saiki R K, Scharaf S, Faloona F, Mullis K B, Horn G T, Erlich HA, Arnheim N. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science* 1985; **230**: 1350-1354.

Salooja N, Catto A, Carter A, Tudenham E G D, Grant P J. Methylene tetrahydrofolate reductase C677T genotype and stroke. *Clinical Laboratory Haematology* 1998; **20**: 357-361.

Sarkar G, Kapelner S, Sommer S S. Formamide can dramatically improve the specificity of PCR. *Nucleic Acids Research* 1990; **18**: 7465.

Sawaya M R, Kraut J. DHFR the movie. *http://chem-faculty.ucsd.edu/kraut/dhfr.html* Visited 17/06/2006.

Schena M, Shalon D, Davis R W, Brown P O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; **270**:467-470.

Schild D, Brake A J, Kiefer M C, Young D, Barr P J. Cloning of three human multifunctional *de novo* purine biosynthetic genes by functional complementation of yeast mutations. *Proceedings of the National Academy of Science* 1990; **87**: 2916-2920.

Schildkraut C I, Marmur J, Doty P. The formation of hybrid DNA molecules and theirm use in studies of DNA homologies. *Journal of Molecular Biology* 1961; **3**: 595-617.

Schimke R T. Gene amplification in cultured cells. *Journal of Biological Chemistry* 1988; **263**: 5989-5992.

Schmitzer A R, Lepine F, Pelletier J N. Combitorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: creating alternative functional configurations. *Protein Engineering, Design and Selection* 2004; **17**: 809-819.

Schnell J R, Dyson H J, Wright P E. Effect of cofactor binding and loop conformation on side chain methyl dynamics in dihydrofolate reductase. *Biochemistry* 2004; **43**: 374-383.

Schollen E. Dequeker E, McQuaid S, Vankeirsbilck B, Michils G, Harvey J, can den Akker E, van Schooten R, Clark Z, Schrooten S. Matthijs G. Diagnostic DHPLC Quality Accurance (DDQA): A collaborative approach to the generation of validate and standardized methods for DHPLC-based mutation screening in Clinical Genetics Laboratories. *Human Mutation* 2005; **25**: 583-592.

Schreier A A, Gruber J. Meeting highlights. Oncogenic human polyomaviruses. *Journal of the National Cancer Institute* 1986; **76**:1255-1258.

Schriml L M, Peterson R J, Gerrard B, Dean M. Use of denaturibg HPLC to map human and murine genes and to validate single-nucleotide polymorphisms. *Biotechniques* 2000; **28**: 740-745.

Schweitzer B I, Srimatkandada S, Gritsman H, Sheridan R, Venkataraghavan R, Bertino J R. Probing the role of two hydrophobic active site residues in the human dihydrofolate reducatase by site-directed mutagenesis. *Journal of Biological Chemistry* 1989; **264**: 20786-20795.

Scionti I, Michelacci F, Pasello M, Hattinger C M, Alberghini M, Manara M C, Bacci G, Ferrari S, Scotlandi K, Picci P, Serra M. Clinical impact of the methotrexate resistance-associated genes C-MYC and dihydrofolate reductase (DHFR) in high-grade osteosarcoma. *Annals of Oncology* 2008; (Epub ahead of print).

Seeger K, Buchwald D, Peter A, Taube T, von Stackelberg A, Schmitt G, Henze G. *TEL-AML1* fusion in relapsed childhood acute lymphoblastic leukaemia. Blood 1999; **94**: 374-376.

Selhub J. Homocysteine metabolism. *Annual Review of Nutrition* 1999; **19**: 217-246.

Serra M, Reverter-Branchat G, Maurici D, Benini S, Shen J-N, Chano T, Hattinger C-M, Manara M-C, Pasello M, Scotlandi K, Picci P. Analysis of dihydrofolate reductase and reduced folate carrier gene status in relation to methotrexate resistance in osteosarcoma cells. *Annals of Oncology* 2004; **15**: 151-160.

Shapiro J T Stannard B S, Felsenfeld G. The binding of small cations to deoxyribonucleic acid. Nucleotide specificity. *Biochemistry* 1969; **8**: 3233-41.

Sharina I G, Zhao R, Wang Y, Babani S, Goldman I D. Mutational analysis of the functional role of conserved arginine and lysine residues in transmembrane domains of the murine reduced folate carrier. *Molecular Pharmacology* 2001; **59**: 1022-1028.

Sharp L, Little J, Schofield A C, Pavildou E, Cotton S C, Miedzybrodzka Z, Baird J O C, Haites N E, Heys S D, Grubb D A. Folate and breast cancer: the role of polymorphisms in methylenetetrahydrofolate reductase (MTHFR). *Cancer Letters* 2002; **181**: 65-71.

Shaw G M, Lammer E J, Zhu H, Baker M W, Neri E, Finnell R H. Maternal periconceptional vitamin use, genetic variation of infant reduced folate carrier (A80G), and risk of spina bifida. *American Journal of Medical Genetics* 2002; **108**: 1-6.

Shaw G M, Zhu H, Lammer E J, Yang W, Finnell R H. Genetic variation of infant reduced folate carrier (A80G) and risk of orofacial and conotruncal heart defects. *American Journal of Epidemiology* 2003; **158**: 747-752.

Sibani S, Leclerc D, Weisberg I S, O'Ferrall E, Watkins D, Artigas C, Rosenblatt D S, Rozen R. Characterization of mutations in severe methylenetetrahydrofolate reductase deficiency reveals an FAD-responsive mutation. *Human Mutation* 2003; **21**: 509-520.

Sierra E E, Goldman I D. Recent advances in the understanding of the mechanism of membrane transport of folates and antifolates. *Seminars in Oncology* 1999; **26, supplement 6**: 11-23.

Sigmond J, Backus H H J, Wouters D, Temmink O H, Janser G, Peters G J. Induction of resistance to the multitargeted antifolate Pemetrexed (ALIMTA) in WrDr human colon cancer cells is associated with thymidylate synthase overexpression. *Biochemical Pharmacology* 2003; **66**: 431-438.

Singer M J, Mesner L D, Friedman C L, Trask B J, Hamlin J L. Amplification of the human dihydrofolate reductase gene via double minutes is initiated by chromosome breaks. *Proceedings of the National Academy of Science* 2000; **97**: 7921-7926.

Skacel N, Menon L G, Mishra P J, Peters R, Banerjee D, Bertino J R, Abali E E. Identification of amino acids required for the functional up-regulation of human dihydrofolate reductase protein in response to antifolate treatment. *Journal of Biological Chemistry* 2005; **280**: 22721-22731.

Skibola C F, Smith M T, Kane E, Roman E, Rollinson S, Cartwright R A, Morgan G. Polymorphisms in the methylenetetrahydrofolate reductase gene are associated with susceptibility to acute leukemia in adults. *Proceedings of the National Academy of Science* 1999; **96**: 12810-12815.

Skibola C F, Smith M T, Hubbard A, Shane B, Roberts A C, Law G R, Rollinson S, Ronman E, Cartwright R A, Morgan G J. Polymorphisms in the thymidylate synthase and serine hydroxymethyltransferase genes and risk of adult acute lymphocytic leukaemia. *Blood* 2002; **99**: 3786-3791.

Skibola C F, Forrest M S, Coppede F, Agana L, Hubbard A, Smith M T, Bracci P M, Holly E A. Polymorphisms and haplotypes in folate-metabolizing genes and risk of non-Hodgkin lymphoma. *Blood* 2004; **104**: 2155-2161.

Sigmund J, Backus H H J, Wouters D, Temmink O H, Jansen G, Peters G J. Induction of resistance to the multitargeted antifolate Pemetrexed (ALIMTA) in WiDr human colon cancer cells is associated with thymidylate synthase overexpression. *Biochemical Pharmacology* 2003; **66**: 431-438.

Smith M, Arthur D, Camitta B, Carrroll A J, Crist W, Gaynon P, Gelber R, Heerema N, Korn E L, Link M, Murphy S, Pui C-H, Pullen J, Reaman G, Sallan S E, Sather H, Shuster J, Simon R, Trigg M, Tubergen D, Uckun F, Ungerleider R. Uniform approach to risk classification and treatment assignment for children with acute lymphoblastic leukaemia. *Journal of Clinical Oncology* 1996; **14**: 18-24.

Smith M A, Ries L A, Gurney J G, Ross J A. Leukemia. In: Ries L A, Smith M A, Gurney J G, et al (Eds): **Cancer incidence and survival among children and adolescents: United States SEER program 1975-1995**. Bethesda, Md: National Cancer Institute, SEER Program, 1999. NIH Publication Number 99-4649.

Sorich M J, Pottier N, Pei D, Yang D, Yang W, Kager L, Stocco G, Cheng C, Panetta J C, Pui C-H, Relling M V, Cheok M H, Evans W E. In vivo response to methotrexate forecasts outcome of acute lymphoblastic leukaemia and has a distinct gene expression profile. *PLoS Medicine* 2008; **5**: e83. DOI: 10.1371/journal.pmed.0050083.

Sowers R, Toguchida J, Qin J, Meyers P A, Healey J H, Huvos A, Banerjee D, Bertino J R, Gorlick R. mRNA expression levels of E2F transcription factors correlate with dihydrofolate reductase, reduced folate carrier, and thymidylate synthase mRNA expression in osteosarcoma. *Molecular Cancer Therapeutics* 2003; **2**: 535-541.

Spencer H T, Sorrentino B P, Pui C-H, Chunduru S K, Sleep S E H, Blakley R L. Mutations in the gene for human dihydrofolate reductase: an unlikely cause of clinical relapse in pediatric leukemia after therapy with methotrexate. *Leukemia* 1996; **10**: 439-446.

Spiegelman J I, Mindrinos M N, Oefner P J. High-accuracy DNA sequence variation screening by DHPLC. *BioTechniques* 2000; **29**: 1084-1092.

Srimitkandada S, Medina W D, Cashmore A R, Whyte W, Engel D, Moroson B A, Franco C T, Dube S K, Bertino J R. Amplification and organization of dihydrofolate reductase genes in a human leukemic cell line, K-562, resistant to methotrexate. *Biochemistry* 1983; **22**: 5774-5781.

Srimitkandada S, Schweitzer B I, Moroson B A, Dube S, Bertino J R. Amplification of a polymorphic dihydrofolate reductase gene expressing an enzyme with decreased binding to methotrexate in a human colon carcinoma cell line, HCT-8R4, resistant to this drug. *Journal of Biological Chemistry* 1989; **264**: 3524-3528.

Stadt U Z, Rischewski J, Schneppenheim R, Kabisch H. Denaturing HPLC for identification of cloncal receptor γ rearrangements in newly diagnosed acute lymphoblastic leukaemia. *Clinical Chemistry* 2001; **47**: 2003-2011.

Stanisławska-Sachadyn A, Brown K S, Mitchell L E, Woodside J V, Young I S, Scott J M, Murray L, Boreham C A, McNulty H, Strain J J, Whitehead A S. An insertion/deletion polymorphism of the dihydrofolate reductase (*DHFR*) gene is associated with serum and red blood cell folate concentrations in women. *Human Genetics* 2008; **123**: 289-295.

Stark M, Hayward N. genome-wide loss of heterozygosity and copy number analysis in melanoma using high-density single-nucleotide polymorphism arrays. *Cancer Research* 2007; **67**: 2632-2642.

Steenge G R, Verhoef P, Katan M B. Betaine supplementaiton lowers plasma homocysteine in healthy men and women. *Journal of Nutrition* (e-version 2002 ahead of printed version in 2003); **133**: 1291-1295.

Stockman B J, Nirmala N R, Wagner G, Delcamp T J, DeYarman M T, Freisheim J H. Sequence-specific ^{1}H and ^{15}N resonance assignments for human dihydrofolate reductase in solution. *Biochemistry* 1992; **31**: 218-229.

Subramanian V S, Chatterjee N, Said H M. Folate uptake in the human intestine promoter activity and effect of folate deficiency. *Journal of Cell Physiology* 2003; **196**: 403-408.

Swinscow T D V. *Statistics at square one*. British Medical Association, London. 1983; 58-61.

Synold T W, Relling M V, Boyett J M, Rivera G K, Sandlund J T, Mahmoud H, Crist W M, Pui C-H, Evans W E. Blast cell methotrexate-polyglutamate accumulation in vivi differes by lineage, ploidy and methotrexate doses in acute lymphoblastic leukemia. *Journal of Clinical Investigation* 1994; **94**: 1996-2001.

Tabone T, Sallmann G, Webb E, Cotton R G H. Dtection of 100% of mutations in 124 individuals using a standard UV/Vis microplate reader: a novel concept for mutation scanning. *Nucleic Acids Research* 2006; **34**: e45 (published on-line).

Tai N, Ding Y, Schmitz J C, Chu E. Identification of critical amino acid residues on human dihydrofolate reductase protein that mediate RNA recognition. *Nucleic Acids Research* 2002; **30**: 4481-4488.

Tai N, Schmitz J C, Chen T-M, Chu E. Characterization of a *cis*-acting regulatory element in the protein-coding region of human dihydrofolate reductase mRNA. *Biochemical Journal* 2004; **378**: 999-1006.

Takeishi K, Kaneda S, Ayusawa D, Shimizu K, Gotoh O, Seno T. Nucleotide sequence of a functional cDNA for human thymidylate synthase. *Nucleic Acids Research* 1985; **13**: 2035-2043.

Tan X, Huang S, Ratnam M, Thompson P D, Freisheim J H. The importance of loop region residues 40-46 in human dihydrofolate reductase as revealed by site-directed mutagenesis. *Journal of Biological Chemistry* 1990; **265**: 8027-8032.

Taub J W, Ge Y. Down Syndrome, drug metabolism, and chromosome 21. *Paediatric Blood Cancer* 2005; **44**: 33-39.

Taub J W, Konrad M A, Ge Y, Naber J M, Scott J S, Matherly L H, Ravindranath Y. High frequency of leukaemic clones in newborn screening blood samples of children with B-precursor acute lymphoblastic leukaemia. *Blood* 2002; **99**: 2992-2996.

Teuffel O, Dettling M, Cario G, Stanulla M, Schrappe M, Buhlmann P, Niggli F K, Schafer B W. Gene expression profiles and risk stratification in childhood acute lymphoblastic leukemia. *Haematolgica* 2004; **89**: 801-808.

Thompson P D, Freisheim J H. Coversion of arginine to lysine at position 70 of human dihydrofolate reductase: generation of a methotrexate-insensitive mutant enzyme. *Biochemistry* 1991; **30**: 8124-8130.

Thompson J R, Fitzgerald P, Willoughby L N, Armstrong B K. Maternal folate supplementation in pregnancy and protection against acute lymphoblastic leukaemia in childhood: a case-control study. *Lancet* 2001; **358**: 1935-1940.

Thorpe I F, Brooks C L. The coupling of structural fluctuation to hydride transfer in dihydrofolate reductase. *Proteins: Structure, Function and Bioinformatics* 2004; **57**: 444-457.

Tibshirani R, Hastie T, Eisen M, Ross D, Bolstein D, Brown P. Clustering methods for the analysis of DNA microarray data. *Stanford University Report* 1999; 1-23.

Tolner B, Roy K, Sirotnak F M. Organization, structure and alternate splicing of the murine RFC-1 gene encoding a folate transporter. *Gene* 1997; **189**: 1-7.

Tolner B, Roy K, Sirotnak F M. Structural analysis of the human RFC-1 gene encoding a folate transporter reveals multiple promoters and alternatively spliced transcripts with 5' end hetergeneity, *Gene* 1998; **211**: 331-341.

Tombline G, Bellizzi D, Sgaramella V. Heterogeneity of primer extension products in asymmetricd PCR is due to both to cleavage by a structure-specific exo/endonuclease activity of DNA polymerases and to premature stops. *Proceedings of the National Academy of Science* 1996; **93**: 2724-2728.

Toole J F, Malinow M R, Chambless L E, Spence J D, Pettigrew L C. Howard V J, Sides E G, Wang C-H, Stampfer M. Lowering homocysteine in patients with ischaemic stroke to prevent recurrent stroke, myocardial infarction, and death. *Journal of the American Medical Association* 2004; **291**: 565-575.

Trippett T, Schlemmer S, Elisseyeff Y, Goker E, Wachter M, Steinherz P, Tan C, Berman E, Wright J E, Rosowsky A, Schweitzer B, Bertino J R. Defective transport as a mechanism of acquired resistance to methotrexate in patients with acute lymphocytic leukemia. *Blood* 1992; **80**: 1158-1162.

Trippett T M, Bertino J R. Therapuetic strategies targeting proteins that regulate folate and reduced folate transport. *Journal of Chemotherapy* 1999; **11**: 3-10.

Trippett T M, Garcia S, Manova K, Mody R, Cohen-Gould L, Flintoff W, Bertino J R. Localization of a human reduced folate carrier protein in the mitochondrial as well as the cell membrane of leukemia cells. *Cancer Research* 2001; **61**: 1941-1947.

Tsay J-T, Appleman J R, Beard W A, Prendergast N J, Delcamp T J, Freisheim J H, Blakley R L. Kinetic investigation of the functional role of phenylalanine-31 of recombinant human dihydrofolate reductase. *Biochemistry* 1990; **29**: 6428-6436.

Tse A, Brigle K, Taylor S M, Moran R G. Mutations in the reduced folate carrier gene which confer dominant resistance to 5,10-dideazatetrahydrofolate. *Journal of Biological Chemistry* 1998; **273**: 25953-25960.

Tsuzuki S, Karnan S, Horibe K, Matsumoto K, Kato K, Inukai T, Goi K, Sugita K, Nakazawa S, Kasugai Y, Ueda R, Seto M. Genetic abnormalities involved in t(12;21) *TEL-AML1* acute lymphoblastic leukaemia: Analysis by means of array-based comparative genomic hybridization. *Cancer Science* 2007; **98**: 698-706.

Underhill P A, Jin L, Lin A A, Mehdi S Q, Jenkins T, Vollrath D, Davis R W, Cavalli-Sforza L L, Oefner P J. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Research* 1997; **7**: 996-1005.

Urano W, Taniguchi A, Yamanaka H, Tanaka E, Nakajima H, Matsuda Y, Akama H, Kitamura Y, Kitatani N. Polymorphisms in the methylenetetrahydrofolate reductase gene were associated with both the efficacy and the toxicity of methotrexate used for the treatment of rheumatoid arthritis, as evidenced by single locus and haplotype analyses. *Pharmacogenetics* 2002; **12**: 183-190.

Urano W, Taniguchi A, Yamanaka H, Tanaka E, Nakajima H, Matsuda Y, Akama H, Kitamura Y, Kitatani N, Kamatini N. The efficiecy and the toxicity of methotrexate in rheumatoid arthritis patients are associated with polymorphims in the methylenetetrahydrofolate reductase gene. **Programme & Abstracts, Joint 11th International and 9th European Symposium on Purines and Pyrimidines in man:** 2003; PP03: 001.

Van't Veer L J, Dai H, van de Vijver M J, He Y D, Hart A A, Mao M, Peterse H L, van der Kooy K, Marton M J, Witteveen A T, Schreiber G J, Kerkhoven R M, Roberts C, Linsley P S, Bernards R, Friend S H. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; **415**: 530-536.

Varadaraj K, Skinner D M. Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases. *Gene* 1994; **140**: 1-5.

Voet D, Voet J G, Pratt C W. *Fundamentals of Biochemistry* (2nd Edition) John Wiley & Sons: New York. 2006: 705.

Volm M, Sauerbrey A, Zintl F. Dihydrofolate-reductase and thymidylate-synthase in childhood acute lymphoblastic leukemia. *Anticancer Research* 1994; **14**: 1377-1382.

Volpato J P, Fossati E, Pelletier J N. Increasing methotrexate resistance by combination of active-site mutations in human dihydrofolate reductase. *Journal of Molecular Biology* 2007; **373**: 599-611.

Wade M, Blake M C, Jambou R C, Helin K, Harlow E, Azizkhan J C. An inverted repeat motif stabilizes binding of E2F and enhances transcription of the dihydrofolate reductase gene. *Journal of Biological Chemistry* 1995; **270**: 9783-9791.

Wagner T, Stoppa-Lyonnet D, Fleischmann E, Muhr D, Pages S, Sandberg T, Caux V, Moeslinger R, Langbaurer G, Borg A, Oefner P. Denaturing high-performance liquid chromatography detects reliably BRCA-1 and BRCA-2 mutations. *Genomics* 1999; **62**: 369-376.

Wall N R, Mohammad R M, Al-Katib A M. Bax:Bcl-2 ratio modulation by brystatin 1 and novel antitubulin agents is important for susceptibility to drug induced aopotosis in the human early pre-B acute lymphoblastic leukemia cell line,Reh. *Leukemia Research* 1999; **23**: 881-888.

Wang D G and Lipshutz R. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998; **280:** 1077-1082.

Wang G, Ahmad KA, Ahmed K. Role of protein kinase CK2 in the regulation of tumor necrosis factor-related apoptosis inducing ligand-induced apoptosis in prostate cancer cells. *Cancer Research* 2006; **66**: 2242-2249.

Wang X, Shen F, Freisheim J H, Gentry L E, Ratnam M. Differential stereospecificities and affinities of folate receptor isoforms for folate compounds and antifolates. *Biochemical Pharmacology* 1992; **44**: 1898-1901.

Wang X, Fenech M. A comparison of folic acid and 5-methyltetrahydrofolate for prevention of DNA damage and cell death in human lymphocytes in vitro. *Mutagenesis* 2003; **18**: 81-86.

Wang Y, Zhao R, Russell R G, Goldman D. Localization of the murine reduced folate carrier as assessed by immunohistochemical analysis. *Biochimica et Biophysica Acta* 2000; **1513**: 49-54.

Wang Y, Zhao R, Goldman I D. Decreased expression of the reduced folate carrier and folylpolyglutamate synthetase is the basis for acquired resistance to the pemetrexed antifolate (LY231514) in an L1210 murine leukemia cell line. *Biochemical Pharmacology* 2003; **65**: 1163-1170.

Wang Y, Zhao R, Goldman I D. Characterization of a folate transporter in HeLa cells with a low pH optimum and high affinity for pemetrexed distinct from the reduced folate carrier. *Clinical Cancer Research* 2004; **10**: 6256-6264.

Wang Y, Rajgopal A, Goldman I D, Zhao R. Preservation of folate transport activity with a low-pH optimum in rat IEC-6 intestinal epithelial cell lines that lack reduced folate carrier function. *Americal Journal Physiology: Cell Physiology* 2005; **288**: C65-C71

Watkins D, Ru M, Hwand H-Y, Kim C D, Murray A, Philip N S, Kim W, Legakis H, Wai T, Hilton J F, Ge B, Dore C, Hosack A, Wilson A, Gravel R A, Shane B, Hudson T J, Rosenblatt D S. Hyperhomocysteinemia due to methionine synthase deficiency, cblG: structure of the MTR gene, genotype diversity, and recognition of a common mutation, P1173L. *American Journal of Human Genetics* 2002; **71**: 143-153.

Watney J B, Agarwal P K, Hammes-Schiffer S. Effect of mutation on enzyme motion in dihydrofolate reductase. *Journal of the American Chemical Society* 2003; **125**: 3745-3750.

Watson J D, Crick F H C. Molecular structure of nucleic acids. *Nature* 1953a; **171**: 737-738.

Watson J D, Crick F H C. Genetical implications of the structure of deoxyribonucleic acid. *Nature* 1953b; **171**: 964-969.

Weinreb D B, Desman G T, Amolat-Apiado M J M, Burstein D E, Godbold J H, Johnson E M. Polyoma virus infection is a prominent risk factor for bladder carcinoma in immunocompetent individuals. *Diagnostic Cytopathology* 2006; **34**: 201-203.

Weissensteiner T, Lanchbury J S. Strategy for controlling preferential amplification and avoiding false negatives in PCR typing. *BioTechniques* 1996; **21**: 1102-8

Weitman S D, Weinberg A G, Coney LR, Zurawski V R, Jennings D S, Kamen B A. Cellular localization of the folate receptor: potential role in drug toxicity and folate homeostasis. *Cancer Research* 1992; **52**: 6708-6711.

Weisberg I, Tran P, Christensen B, Sibani S, Rozen R. A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity. *Molecular Genetics and Metabolism* 1998; **64**: 16-172.

Westerhof G R, Jansen G, van Emmerik N, Kathmann I, Rijksen G, Jackman A L, Schornagel J H. Membrane transport of natural folates and antifolate compounds in murine L1210 leukemia cells: role of carrier- and receptor-mediated transport systems. *Cancer Research* 1991; **51**: 5507-5513.

Whetstine J R, Matherly L H. The basal promoters for the human reduced folate carrier gene are regulated by a GC-box and a cAMP-response element/AP-1-like element. *Journal of Biological Chemistry* 2001a; **276**: 6350-6358.

Whetstine J R, Gifford A J, Witt T, Liu X Y, Flately R M, Norris M, Haber M, Taub J W, Ravindranath Y, Matherly L H. Single nucleotide polymorphisms in the human reduced folate carrier: characterization of a high-frequency G/A variant at position 80 and transport properties of the His^{27} and Arg^{27} carriers. *Clinical Cancer Research* 2001b; **7**: 3416-3422.

Whetstine J R, Flatley R M, Matherly L H. The human reduced folate carrier gene is ubiquitously and differentially expressed in normal human tissues: identification of seven non-coding exons and characterization of a novel promoter. *Biochemical Journal* 2002b; **367**: 629-640.

Whetstine J R, Witt T L, Matherly L H. The human reduced folate carrier gene is regulated by the AP2 and Sp1 transcription factor families and a functional 61 base pair polymorphism. *Journal of Biological Chemistry* 2002a; **277**: 43873-43880.

Whitehead V M, Vuchich M J, Lauer S J, Mahoney D, Carroll A J, Shuster J J, Essletine D W, Payment C, Look A T, Akabutu J, Bowen T, Taylor L D, Camitta B, Pullen D J. Accumulation of high levels of methotrexate polyglutmates in lymphoblsts from children with hyperdiploid (>50 chromosomes) B-lineage acute lymphoblastic leukemia: A Pediatric Oncology Group study. *Blood* 1992; **80**: 1316-1323.

Whitehead V M, Payment C, Cooley L, Lauer S J, Mahoney D H, Shuster J, Vuchich M-J, Bernstein M L, Look A T, Pullen D J, Camitta B. The association of the TEL-AMPL1 chromosomal translocation with the accumulation of methotrexate polyglutamates in lymphoblasts and with ploidy in childhood B-progenitor cell acute lymphoblastic leukemia: a Pediatric Oncology Group study. *Leukemia* 2001; **15**: 1081-1088.

Wickramasinghe S N and Frida S. Bone marrow cells from vitamin B_{12} and folate-deficient patients misincorporate uracil into DNA. *Blood* 1994; **83**: 1656-1661.

Wiemels J L, Smith R N. Taylor S G, Eden O B, Alexander F E, Greaves M F and United Kingdom Childhood Cancer Study Investigators. Methylenetetrahydrofolate reductase (MTHFR) polymorphisms and risk of molecularly defined subtypes of childhood acute leukemia. *Proceedings of the National Academy of Science* 2001; **98**: 4004-4009.

Williams F M R, Murray R C, Underhill M, Flintoff W F. Isolation of a hamster cDNA clone coding for a function involved in methotrexate uptake. *Journal of Biological Chemistry* 1994; **269**: 5810-5816.

Williams F M R, Flintoff W F. Isolation of a human cDNA that complements a mutant hamster cell defective in methotrexate uptake. *Journal of Biological Chemistry* 1995; **270**: 2987-2992.

Witt T L, Matherly L H. Identification of lysine-411 in the human reduced folate carrier as an important determinant of substrate selectivity and carrier function by systematic site-directed mutagenesis. *Biochimica et Biophysica Acta* 2002; **1567**: 56-62.

Witt T L, Stapels S E, Matherly L H. Restoration of transport activity by co-expression of human reduced folate carrier half-molecules in transport-impaired K562 cells. *Journal of Biological Chemistry* 2004; **279**: 46755-46763.

Wong S C, Proefke S A, Bhushan A, Matherly L H. Isolation of human cDNAs that restore methotrexate sensitivity and reduced folate carrier activity in methotrexate transport-defective chinese hamster ovary cells. *Journal of Biological Chemistry* 1995; **270**: 17468-17475.

Wong S C, Zhang L, Proefke S A, Matherly L H. Effects of the loss of capacity for N-glycoslylation on the transport activity and cellular localization of the human reduced folate carrier. *Biochimica et Biophysica Acta* 1998; **1375**: 6-12.

Wong S C, Zhang L, Witt T K, Proefke S A, Bhushan A, Matherly L H. Impaired membrane transport in methotrexate-resistant CCRF-CEM cells involves early translation termination and increased turnover of a mutant reduced folate carrier. *Journal of Biological Chemistry* 1999; **274**: 10388-10394.

Worm J, Kirken A F, Dzhandzhugazyan K N. Guldberg P. Methylation-depended silencing of the reduced folate carrier gene inherently methotrexate-resistant human breast cancer cells. *Journal of Biological Chemistry* 2001; **276**: 39990-40000

Wright G, Tan B, Rosenwald A, Hurt E H, Wiestner A, Staudt L M. A gene experession-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proceedings of the National Academy of Science* 2003; **100**: 9991-9996.

Xiao W, Oefner P J. Denaturing high-performance liquid chromatography: a review. *Human Mutation* 2001; **17**: 439-474.

Xie Y, Davies S M, Xiang Y, Robison L L, Ross J A. Trends in leukaemia incidence and survival in the United States (1973-1998). *Cancer* 2003; **97**: 2229-2235.

Xu X, Gammon M D, Wetmur J G, RAo M, GAudet M M, Teitelbaum S L, Britton J A, Neugut A I, Santella R M, Chen J. A functional 19-base pair deletion polymorphism of dihydrofolate reductase (DHFR) and risk of breast cancer in multivitamin users. *American Journal of Clinical Nutrition* 2007; **85**: 1098-1102.

Yamada K, Chen Z, Rozen R, Matthews R G. Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase. *Proceedings of the National Academy of Science* 2001; **98**: 14853-14858.

Yang R, Sowers R, Mazza B A, Healey J H, Huvos A, Grier H, Bernstein M, Beardsley P, Krailo M D, Devidas M, Bertino J R, Meyesr P A, Gorlick R. Sequence alterations in the reduced folate carrier are observed in osteosarcoma tumor samples. *Clinical Cancer Research* 2003; **9**: 837-844.

Yang X, Davies S M, Xiang Y, Robison L L, Ross J A. Trends in leukemia incidence and survival in the United States (1973-1998). *Cancer* 2003; **97**: 2229-2235.

Yang-Feng T L, MA Y-Y, Liang R, Prasad P D, Leibach F H, Ganapathy V. Assignment of the human folate transporter gene to chromosome 21q22.3 by somatic cell hybrid analysis and in situ hybridisation. *Biochemical and Biophysical Research Communications* 1995; **210**: 874-879.

Yates Z, Lucock M. G80A reduced folate carrier SNP modulates cellular uptake of folate and affords protection against thrombosis via a non-homocysteine related mechanism. *Life Sciences* 2005; **77**: 2735-2742.

Yeoh E-J, Ross M E, Shurtleff S A, Williams W K, Patel D, Mahfouz R, Behm F G, Raimondi S C, Relling M V, Patel A, Cheng C, Campana D, Wilkins D, Zhou X, Li J, Liu H, Pui C-H, Evans W E, Naeve C, Wong L, Downing J R. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemai by gene expression profiling. *Cancer Cell* 2002; **1**: 133-143.

Yi P, Pogribny I P, James S J. Multiplex PCR for simultaneous detection of 677 C→T and 1298 A→C polymorphisms in methylenetetrahydrofolate reductase gene for population studies of cancer risk. *Cancer Letters* 2002; **181**: 209-213.

Yin D, Chave K J, Macaluso C R, Galivan J, Yao R. Structural organization of the human γ-glutamy hydrolase gene. *Gene* 1999; **238**: 463-470.

Zeng H, Liu G, Rea P A, Kruh G D. Transport of amphipathic anions by human multidrug resistance protein 3. *Cancer Research* 2000; **60**: 4779-4784.

Zeng H, Chem Z-S, Belinsky M G, Rea P A, Kruh G D. Transport of methotrexate (MTX) and folates by multidrug resistance protein (MRP) 3 and MRP1: Effect of polyglutamylation on MTX transport. *Cancer Research* 2001; **61**: 7225-7232.

Zhang L, Wong S C, Matherly L H. Transcript heterogeneity of the human reduced folate carrier results from the use of multiple promoters and variable splicing of alternative upstream exons. *Biochemical Journal* 1998a; **332**: 773-780.

Zhang L, Wong S C, Matherly L H. Structure and organization of the human reduced folate carrier gene. *Biochemica et Biophysica Acta* 1998b; **1442**: 389-393.

Zhang L, Taub J W, Williamson M, Wong, S C, Hukku B, Pullen J, Ravindranath Y, Matherly L H. Reduced folate carrier gene expression in childhood acute lymphoblastic leukemia: Relationship to immunophenotype and ploidy. *Clinical Cancer Research* 1998c; **4**: 2169-2177.

Zhao R, Assaraf Y G, Goldmann I D. A mutated murine reduced folate carrier (RFC1) with increased affinity for folic acid, decreased affinity for methotrexate, and an obligatory anion requirement for transport function. *Journal of Biological Chemistry* 1998a; **273**: 19065-19071.

Zhao R, Assaraf Y G, Goldmann I D. A reduced folate carrier mutation produces substrate-dependent alterations in carrier mobility in murine leukemia cells and methotrexate resistance with conservation of growth in 5-formyltetrahydrofolate. *Journal of Biological Chemistry* 1998b; **273**: 7873-7879.

Zhao R, Sharina I G, Goldmann I D. Pattern of mutations that results in loss of reduced folate carrier function under antifolate selective pressure augmented by chemical mutagenesis. *Molecular Pharmacology* 1999; **56**: 68-76.

Zhao R, Gao F, Wang P J, Goldmann I D. Role of the amino acid 45 residue in reduced folate carrier function and ion-dependent transport as characterized by site-directed mutagenesis. *Molecular Pharmacology* 2000; **57**: 317-323.

Zhao X, Li C, Mok S C, Chen Z, Wong D T W. Whole genome loss of heterozygosity profiling on oral squamous cell carcinoma bu high-density single nucleotide polymorphic allele (SNP) array. *Cancer Genetics and Cytogenetics* (2003 e-publication ahead of printed version in 2004); **151**: 82-84

Zhao R, Hanscom M, Goldman I D. The relationship between folate transport activity at low pH and reduced folate carrier function in human Huh7 hepatoma cells. *Biochimie et Biophysica Acta* 2005; **1715**: 57-64.

Zheng H-T, Peng Z-H, Li S, He L. Loss of heterozygosity analyzed by single nucleotide polymorphism array in cancer. *World Journal of Gastroenterology* 2005; **11**: 6740-6744.

Zhou X, Mok S C, Chen Z, Li Y, Wong D T W. Concurrent analysis of loss of heterozygosity (LOH) and copy number abnormality (CAN) for oral premalignancy progression using the Affymetrix 10K SNP mapping array. *Human Genetics* 2004; **115**: 327-330.

Zhou X, Rao N P, Cole S, Mok S C, Chen Z, Wong D T. Progress in concurrent analysis of loss of heterozygosity and comparative genomic hybridization utilizig high density sigle nucleotide polymorphism arrays. *Cancer Genetics and Cytogenetics* 2005; **159**: 53-57.

Zipursky A, Brown E, Christensen H, Sutherland R, Doyle J. Leukemia and/or myeloproliferative syndrome in neonates with Down Syndrome. *Seminars in Perinatology* 1997; **21**: 97-101.

Zuna J, Ford A M, Peham M, Patel N, Saha V, Eckert C, Kochling J, Panzer-Grumayer R, Trka J, Greaves M. *TEL* deletion analysis supports a novel view of relapse in childhood acute lymphoblastic leukemia. *Clinical Cancer Research* 2004; **10**: 5355-5360.

APPENDIX 1

A1 List of equipment used

ABI Prism 377 DNA Sequencer; Applied Biosystems; Hemel Hempstead, UK

BioRad Gel Doc 2000 system; BioRad Laboratories Ltd, Hemel Hempstead, UK

DNASep® DHPLC column; Transgenomic Inc., San Jose, CA

Eppendorf Mastercycler®; Eppendorf, Cambridge, UK

GeneChip® 50K SNP Microarray; Affymetrix Inc, Santa Clara, CA

Joan A14 microcentrifuge

NAVIGATOR™ integrated software; Transgenomic Inc., San Jose, CA

UV microtitre plate reader; Costar, High Wycombe, UK

WAVE® Nucleic Acid Fragment Analyser; Transgenomic Inc., San Jose, CA

WAVE®HS DHPLC system; Transgenomic Inc., San Jose, CA

WAVEMAKER® integrated software; Transgenomic Inc., San Jose, CA

A2 List of reagents used

Agarose (MB1202); Melford Laboratories Ltd, Ipswich, UK.

AmpliTaq Gold® DNA polymerase (5 U/μL); Applied Biosystems, Foster City, CA.

Betaine (anhydrous); Sigma-Aldrich, Poole, UK.

dNTP Master Mix (100 mmol/L); Bioline, London, UK.

Dimethyl sulphoxide (DMSO); Sigma-Aldrich, Poole, UK.

DNA ladders:-

G3161DNA ladder; Promega, Southampton, UK .

P9577 DNA ladder; Sigma, Poole, UK.

100 bp ladder (Product number 15628-019); Invitrogen Ltd, Paisley, UK.

Ethidium bromide(10 mg/mL); Sigma-Aldrich, Poole, UK.

Loading dye G1881 6x concentrate; Promega, Southampton, UK.

$MgCl_2$ (25 mmol/L); Applied Biosystems, Foster City, CA.

$MgSO_4$ (15 mmol/L); Transgenomic Inc, San Jose, CA .

Optimase™ DNA polymerase; Transgenomic Inc, San Jose, CA.

QIAquick PCR Purification Kit (Product code 28104); QIAGEN, Crawley, UK.

QIAamp DNA minicolumn extraction kit (product code 51104); QIAGEN Ltd, Crawley, UK.

Tris-HCl buffer (100 mmol/L Tris-HCl, pH 8.3, 500 mmol/L KCl) ; Applied Biosystems, Foster City, CA.

A3 Final list of primers used

RFC Ex2 forward	CTG ACT CCA CCC CTC CTT CCA GGC ACA GTG
RFC Ex2 reverse	CCA CAT GCC TGC TCC CGC GTG AAC TTC T
RFC Ex3a forward	GTT CAG AGA GGA CGA TGC AG
RFC Ex3a reverse	GAA GGT GAG GAA GGC CAG
RFC Ex t3.2 forward	TCT CCT TCT CCA CGC TCA AC
RFC Ex t3.2 reverse	AAG AAG CCT CGG GGA CC
RFC Ex4 forward	CCC CTC TCA GGC GCC ATC ACG TCC TTC
RFC Ex4 reverse	GTG GAG GGC CTG GGG GAG CAG CAA
RFC Ex5 forward	CTT ATT CTC CCA GCA GCG CC
RFC Ex5 reverse	GCC TCA ACA ATG TCC CCA CA
RFC p6 forward	TGA CCG CTG TGC AGT GTT CAG
RFC p6 reverse	CGC CTG CAA AGT TAC CAC AGG
RFC mRNA forward	GAA ACT CCT GTC CTG GGG AG
RFC mRNA reverse	GCG CCA TGA AGC CGT AGA AG
DHFR Exon 2m for	CGC CTG CAC AAA TGG GGA CGA G
DHFR Exon 2m rev	CCG CGC AGC AGA AAA GGG GAA TC
DHFR Exon 3 for	CCC AGC CCT GGA GAA AAC AC
DHFR Exon 3 rev	ACC CAG CTG CCA ATT CTG CC
DHFR Exon 4 for	AGA CTC CAC ACA GAC GGT GG
DHFR Exon 4 rev	AGG GTT GGG TCC AGA AAG GG
DHFR Exon 5 for	CTG ATG TTA AGT GCT TTT TGT TGA
DHFR Exon 5 rev	TTA TAC CTG TTT CTT CCA CTT CCT
DHFR Exon 6 for	GGC AGC ACC AAG CAT ATT TT
DHFR Exon 6 rev	GCA CCC ATC ATC CTA GCA GT
DHFR Exon 7 for	CTG AGA ATC AGG GAA GCT GTG
DHFR Exon 7 rev	GAC TCA GTT GGG GTC TTG GA

A4 Preparation of working primer solutions

Primers supplied as vials of lyophilised material were reconstituted to a suitable concentration for use in PCR. Each vial was supplied with a data sheet confirming the oligonucleotide sequence, absorbance reading at 260 nm and the amount present in the vial. Given the molecular weight of the primer, the molar concentration was calculated for the solution when 1 mL of deionised and sterile water was added (stock solution).

eg., if molecular weight is 7260 Daltons then;

1 litre of a 1 molar solution contains 7,260 g

1 mL of a 1 molar solution contains 7,260 mg

If the amount provided in the vial is 205.9 µg and this is reconstituted in 1 mL this gives a concentration of 205.9 ug/mL (0.2059 mg/mL) and allows the molar concentration to be determined.

Molar concentration of the reconstituted primer = 0.2059/7260 = 0.00002836 M

= 0.028 mM

= 28 µM

For use, 100 µL of a 5 µM working solution was prepared from the stock as described in the following example;

$$\textit{volume of stock solution } \mu L = \frac{\textit{total volume required } \mu L}{(\textit{original concentration } \mu M \,/\, \textit{required concentration } \mu M)}$$

$$= \frac{100}{(28/5)}$$

$$= 17.9 \ \mu L$$

Therefore to prepare 100 µL of working primer solution for use in the PCR mixture add 17.9 µL of stock solution to 82.1 µL of sterile distilled water and mix.

Appendix 2.1. Part of chromosome 21 amplifed by RFC exon 2 primers (showing primers, coding sequence and key features).

Key:- Exon 2 shown in capital letters. Primers shown as shaded blocks. Other key features are shown in red.
Numbers across the top show base number using the first base in the start codon (blue) as position 1.

-70 -60 -50 -40 -30 -20 -10
ctgactccacc cctccttcca gGCACAGCGT CACCTTCGTC CCCTCCGGAG CTGCACGTGG CCTGAGCAGG

C-43T
SNP rs3177999

C-37A mutation in
relapse 6250

1 11 21 31 41 51 61
ATGGTGCCCT CCAGCCCAGC GGTGGAGAAG CAGGTGCCCG TGGAACCTGG GCCTGACCCC GAGCTCCGGT

Start codon

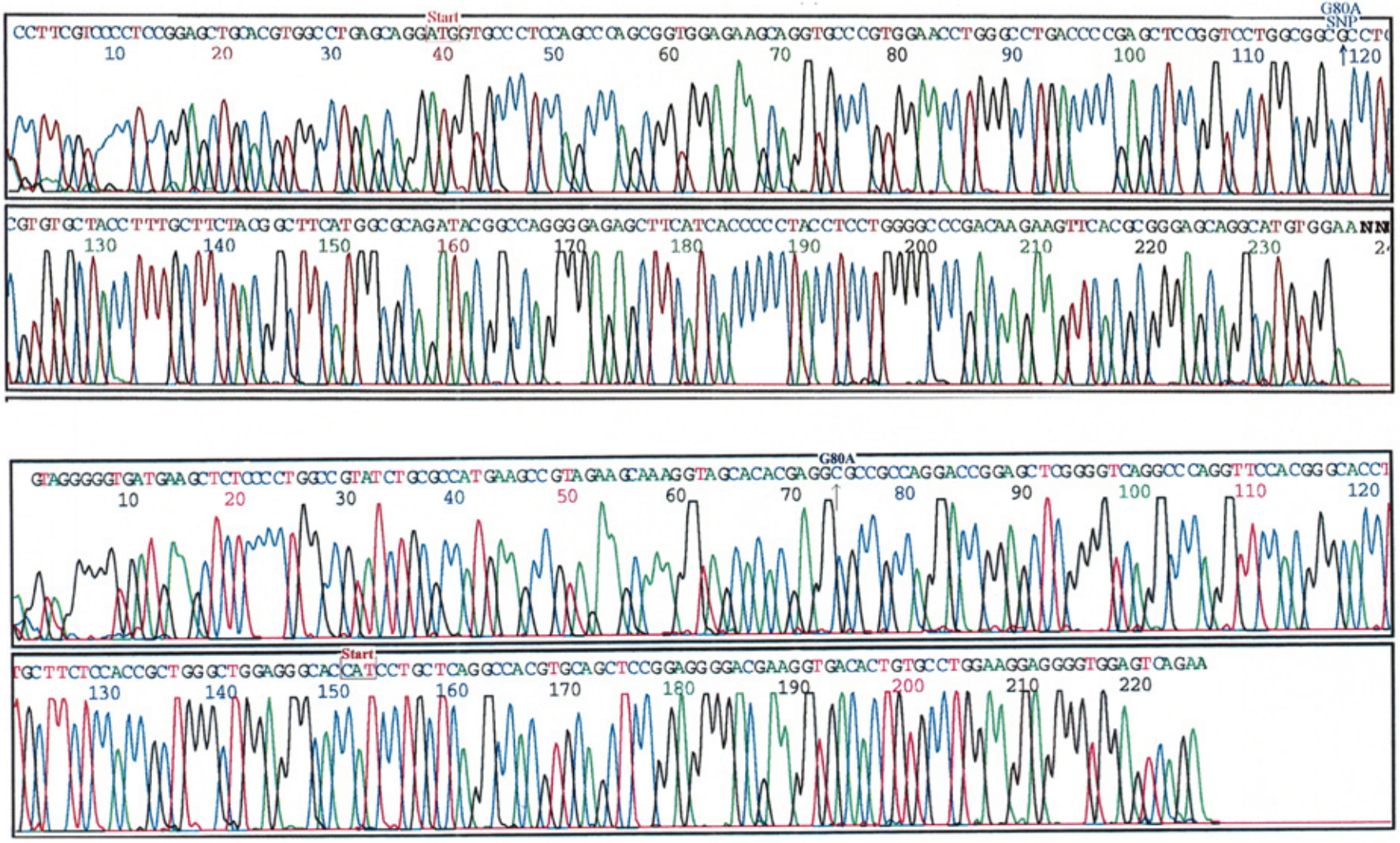
71 81 91 101 111 121 131
CCTGGCGGCA CCTCGTGTGC TACCTTTGCT TCTACGGCTT CATGGCGCAG ATACGGCCAG GGGAGAGCTT

G80A
SNP rs1051266

Glu_{45} CEM/MTX
G133A mutation

141 151 161 171 181 191
CATCACCCCC TACCTCCTGG GGCCCGACAA GAACTTCACG CGGGAGCAGg catgtgg

Appendix 2.2. Sequencing results from RFC exon 2 forward primer (top) and reverse primer (bottom) for NALM6 DNA. Key features are shown; ATG start codon (red), position of known G80A polymorphism (blue).



Appendix 2.3 DHPLC results for RFC exon 2 at 64, 65 and 66°C. Complete profiles at all three temperatures shown in green, samples which have enough data for genotyping of G80A polymorphism shown in yellow.

REDUCED FOLATE CARRIER **268bp**

	DHPLC at 64°C			DHPLC at 65°C			DHPLC at 66°C			
	Original	Spiked		Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
76	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
171	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
279	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	08/07/2003
405	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	08/07/2003
L308	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
980	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
1399	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
1533	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	08/07/2003
1566	?heteroduplex			heteroduplex		G/A				
1720	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
1775	heteroduplex		W/T	heteroduplex		G/A				
1785	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	08/07/2003
2045	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G				
2416	homoduplex	homoduplex	W/T	homoduplex	heteroduplex	A/A	homoduplex	heteroduplex	A/A	08/07/2003
2426	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
2578	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
2650	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
2865	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
3192	?heteroduplex		?	heteroduplex		G/A				
3451	homoduplex	homoduplex	W/T	homoduplex	heteroduplex	A/A	homoduplex	heteroduplex	A/A	08/07/2003
3605	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
3990	homoduplex	homoduplex	W/T	homoduplex	heteroduplex	A/A	homoduplex	heteroduplex	A/A	
4225				homoduplex	heteroduplex	A/A				
4684	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
4845	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	08/07/2003
4900	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	

Continued	DHPLC at 64°C			DHPLC at 65°C			DHPLC at 66°C			
	Original	Spiked		Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
5467				heteroduplex		G/A	heteroduplex		G/A	
5836	homoduplex	heteroduplex	A/A	homoduplex	heteroduplex	A/A				
5862(27/9)	homoduplex	fail		heteroduplex		G/A	heteroduplex		G/A	
5887	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
6250	heteroduplex		G80G C-37A	heteroduplex		G80G C-37A	heteroduplex		G80G C-37A	
6346	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G				
6563	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
6652	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
7096	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
8646	?heteroduplex			heteroduplex		G/A				
8668	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
8678	?heteroduplex			heteroduplex		G/A	heteroduplex		G/A	
8679	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
8839	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
8871			W/T	homoduplex	homoduplex	G/G				
9216	homoduplex	homoduplex	W/T	heteroduplex		G/A				
9506	homoduplex	fail		homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
11420	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N1				heteroduplex		G/A				
N2				heteroduplex		G/A	heteroduplex		G/A	
N3				heteroduplex		G/A	heteroduplex		G/A	
N4				homoduplex	homoduplex	G/G				
N5				homoduplex	homoduplex	G/G				
N6				heteroduplex		G/A				
N7				homoduplex	homoduplex	G/G				
N8				homoduplex	homoduplex	G/G				
N9				heteroduplex		G/A	heteroduplex		G/A	
N11	homoduplex	homoduplex	W/T	heteroduplex		G/A				
N12				homoduplex	homoduplex	G/G				

Appendix 2

Continued	DHPLC at 64°C			DHPLC at 65°C			DHPLC at 66°C			
	Original	Spiked		Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
N13				homoduplex	homoduplex	G/G				
N14				homoduplex	heteroduplex	A/A				
N15				homoduplex	homoduplex	G/G				
N16				homoduplex	heteroduplex	A/A				
N17				homoduplex	heteroduplex	A/A				
N18				heteroduplex		G/A	heteroduplex		G/A	
N19				homoduplex	heteroduplex	A/A				
N20				heteroduplex		G/A	heteroduplex		G/A	
N21				homoduplex	heteroduplex	A/A				
N22				heteroduplex		G/A	heteroduplex		G/A	
N23				heteroduplex		G/A	heteroduplex		G/A	
N24	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
N25	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
N26	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N27				heteroduplex		G/A	heteroduplex		G/A	
N28				heteroduplex		G/A				
N29	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
N30				heteroduplex		G/A	heteroduplex		G/A	
N31				homoduplex	homoduplex	G/G				
N32	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N33				heteroduplex		G/A	heteroduplex		G/A	
N34				homoduplex	homoduplex	G/G				
N35				homoduplex	heteroduplex	A/A	homoduplex	heteroduplex	A/A	
N36	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G				
N37				heteroduplex		G/A	heteroduplex		G/A	
N38	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G				
N39				heteroduplex		G/A	heteroduplex		G/A	
N40				homoduplex	homoduplex	G/G				
N41				homoduplex	homoduplex	G/G				
N42				heteroduplex		G/A	heteroduplex		G/A	

Continued	DHPLC at 64°C			DHPLC at 65°C			DHPLC at 66°C			
	Original	Spiked		Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
N43	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N44	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G				
N45	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N46	homoduplex	homoduplex	W/T	heteroduplex		G/A				
N47				heteroduplex		G/A	heteroduplex		G/A	
N48	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N49				heteroduplex		G/A	heteroduplex		G/A	
N50				homoduplex	homoduplex	G/G				
N51				homoduplex	homoduplex	G/G				
N52				heteroduplex		G/A				
N53	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N54				heteroduplex		G/A	heteroduplex		G/A	
N55	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G				
N56				homoduplex	homoduplex	G/G				
N57	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G				
N58				heteroduplex		G/A	heteroduplex		G/A	
N59	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G				
N60	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N61	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
N62				heteroduplex		G/A				
N63	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
N64	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N65				homoduplex	homoduplex	G/G				
N66	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N67	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N68	homoduplex	homoduplex	W/T	homoduplex	heteroduplex	A/A	homoduplex	heteroduplex	A/A	
N69	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N70	homoduplex	homoduplex	W/T	homoduplex	heteroduplex	A/A				
N71				homoduplex	homoduplex	G/G				
N72				homoduplex	heteroduplex	A/A				

Appendix 2

Continued	DHPLC at 64°C			DHPLC at 65°C			DHPLC at 66°C			
	Original	Spiked		Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
N73	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G	homoduplex	homoduplex	G/G	
N74				heteroduplex		G/A	heteroduplex		G/A	
N75	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N76	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G				
N77	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N78				homoduplex	homoduplex	G/G				
N79	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N80	homoduplex	homoduplex	W/T	heteroduplex		G/A	heteroduplex		G/A	
N81	homoduplex	homoduplex	W/T	homoduplex	homoduplex	G/G				
N83	homoduplex	heteroduplex	A/A	homoduplex	heteroduplex	A/A				

Appendix 2.4 χ^2-test to determine if the genotype distribution is significantly different between the relapse and normal groups.

	Observed				Expected			
Genotype	G/G	G/A	A/A	Total	G/G	G/A	A/A	Total
Normal	31	40	10	81	29.16	42.12	9.72	81
Relapse	14	25	5	44	15.84	22.88	5.28	44
Total	45	65	15	125	45.00	65.00	15.00	125

$\chi^2 = 0.656$. At 2 degrees of freedom this value is $P>0.5$ and therefore hypothesis (null hypothesis of no difference between observed and expected) cannot be rejected.

Therefore there is no significant difference between distribution of genotypes between the normal and relapse data.

Appendix 2.5 Comparison of G80A SNP frequency from this study (final rows) with other published data.

n=			G80G	G80A	A80A	
4806		all data	28.48 (17.6-39.2)	47.74 (33.3-58.0)	23.78 (5.9-40.0)	
3823		all control	27.73 (17.6-33.6)	48.25 (41.4-58.0)	24.02 (10.0-33.3)	
1073		all diseased	29.99 (18.5-39.2)	48.25 (41.0-55.2)	21.76 (5.9-37.0)	
105		total in study	18.1	46.7	35.2	Whetstine 2001b
	54	ALL	18.5	44.5	37.0	
	51	non-ALL	17.6	49.1	33.3	
377		total in study	28.54	49.48	21.98	De Marco 2001
	144	NTD	31.9	50.7	17.4	
	43	Mothers	34.9	48.8	16.3	
	53	Fathers	24.5	51.0	24.5	
	137	Control	22.7	47.4	29.9	
673		total in study	28.7	44.7	26.7	Shaw 2002
	200	NTD	28.1	44.7	27.2	
	473	Control	29.2	44.7	26.1	
462		total in study	27.125	48.175	26.7	Rady 2001
	122	Ashkenazi Jews	28.7	45.9	25.4	
	101	Afro-American	20.8	45.5	33.7	
	131	Caucasians	29.0	47.3	23.7	
	108	Hispanic	26.0	43.8	30.2	
143		total in study	22.5	47.5	30.0	Morin 2003
	56	NTD mothers	23.0	41.0	36.0	
	87	Control	22.0	54.0	24.0	
53		Childhood ALL	30.2	45.3	24.5	Kishi 2003
	9	Seizures	33.3	33.3	33.3	
	5	Thrombosis	20.0	40.0	40.0	
412		total in study	30.1	48.3	21.6	Shaw 2003
	163	Congenital defects	28.8	55.2	16.0	
	249	Control	31.4	41.4	27.2	
273		total in study	32.8	52.0	15.2	Hiraoka 2004
	250	Young Japanes females	33.6	46.0	20.4	
	23	Middle-aged females	32.0	58.0	10.0	
720		total in study	27.375	51.52	21.02	Pei (2004)
		Northern Chinese	31.09	46.63	22.28	
		Southern Chinese	24.75	56.69	18.56	
		Male	25.85	49.27	24.88	
		Female	27.77	53.4	18.83	
	108	Rheumatoid arthritis	31.5	49.1	19.4	Dervieux 2004
	159		33.4	46.5	20.1	Chango 2004

Continued over

152		total in study	34.95	49.75	15.3	Yates 2005
	51	Thrombotic	39.2	54.9	5.9	
	101	Control	30.7	44.6	24.7	
1070		total in study	32.85	49.15	18	Relton 2005
	493	Mother	32.2	47.5	20.3	
	577	Infant	33.5	50.8	15.7	
85		ALL group	35.3	50.6	14.1	de Jonge 2003

Current study

125	both groups	38.3	49.4	12.3
493	Normal	36.0	52.0	12.0
577	ALL	31.8	56.8	11.4

Appendix 3.1. Part of chromosome 21 amplified by RFC exon 3 primers (showing primers, coding sequence and key features).
Key:- Exon 3 shown in capital letters. Primers shown as shaded blocks. Other key features are shown in red.
Numbers across the top show coding base number using the first base in the start codon of exon 2 as position 1.

RFC exon 3.1 forward primer →
gtt cagagaggac gatgcagtgg gacatgggac aggggcgtgt

tggggagagg ggtcccacca ggcagtgttg gggcaggaag ccgtcagcag ccatcacctc gggcgggctg tgtctgccgc agcggcagca ttgctaacac ctggtggggc gcacgcccgc

G/A found in CEM/MTX

190 200 210 220 230 240 250
ctgtccgcag GTCACGAACG AGATCACGCC GGTGCTGTCG TACTCCTACC TGGCCGTGCT GGTGCCGGTG TTCCTGCTCA

G246C SNP rs1051269 Proposed C258T CEM/MTX mutation

260 270 280 290 300 310 320 330
CCGACTACCT GCGCTACACG CCGGTGCTGC TGCTGCAGGG GCTCAGCTTC GTGTCGGTGT GGCTGCTGCT GCTGCTGGGC

340 350 360 370 380 390 400 410
CACTCGGTGG CGCACATGCA GCTCATGGAG CTCTTCTACA GCGTCACCAT GGCCGCGCGC ATCGCCTATT CCTCCTACAT

420 430 440 450 460 470 480 490
CTTCTCTCTC GTGCGGCCCG CGCGCTACCA GCGTGTGGCC GGCTACTCGC GCGCTGCGGT GCTGCTGGGC GTGTTCACCA

500 510 520 530 RFC exon t3.2 forward primer → 560 ← RFC exon 3.1 reverse
GCTCCGTGCT GGGCCAGCTG CTGGTCACTG TGGGCCGAGT CTCCTTCTCC ACGCTCAACT ACATCTCGCT GGCCTTCCTC

primer 590 600 610 620 630 640 650
ACCTTCAGCG TGGTCCTCGC CCTCTTCCTG AAGCGCCCCA AGCGCAGCCT CTTCTTCAAC CGCGACGACC GGGGGCGGTG

660 670 680 690 700 710 720 730
CGAAACCTCG GCTTCGGAGC TGGAGCGCAT GAATCCCGGC CCAGGCGGGA AGCTGGGACA CGCCCTGCGG GTGGCCTGTG

C696G SNP rs12659

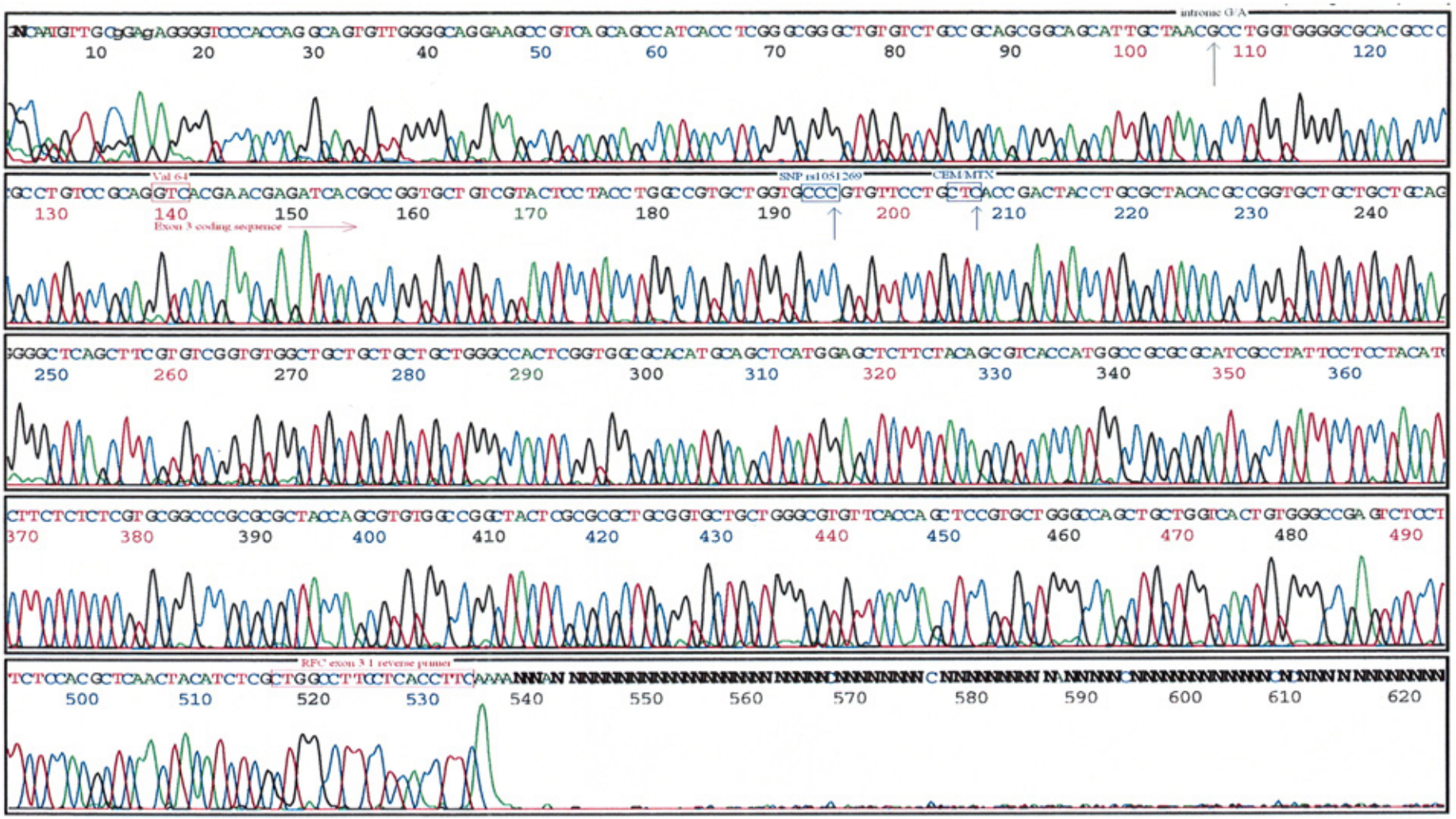
740 750 760 770 780 790 800 810
GGGACTCAGT GCTGGCGCGG ATGCTGCGGG AGCTGGGGGA CAGCCTGCGG CGGCCGCAGC TGCGCCTGTG GTCCCTCTGG

C785T SNP rs928253

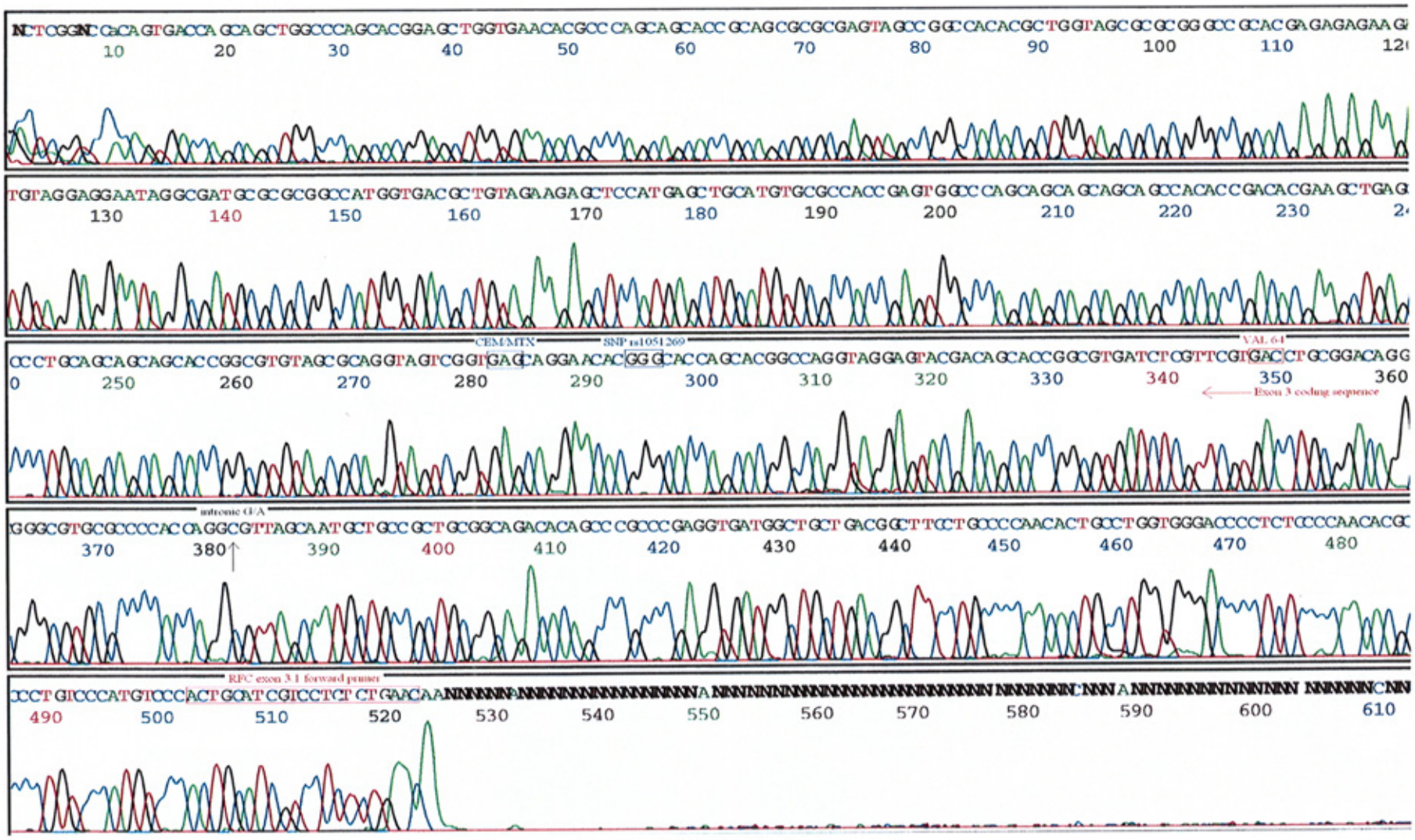
820 830 840 850 860 870 880 890
TGGGTCTTCA ACTCGGCCGG CTACTACCTG GTGGTCTACT ACGTGCACATC CTGTGGAACG AGGTGGACCC CACCACCAAC

900 910 920 930 940 948 ← RFC exon t3.2 reverse primer
AGTGCGCGGG TCTACAACGG CGCGGCAGAT GCTGCCTCCA CGCTGCTGgg tacgcatgcc ctggtccccg aggcttctt

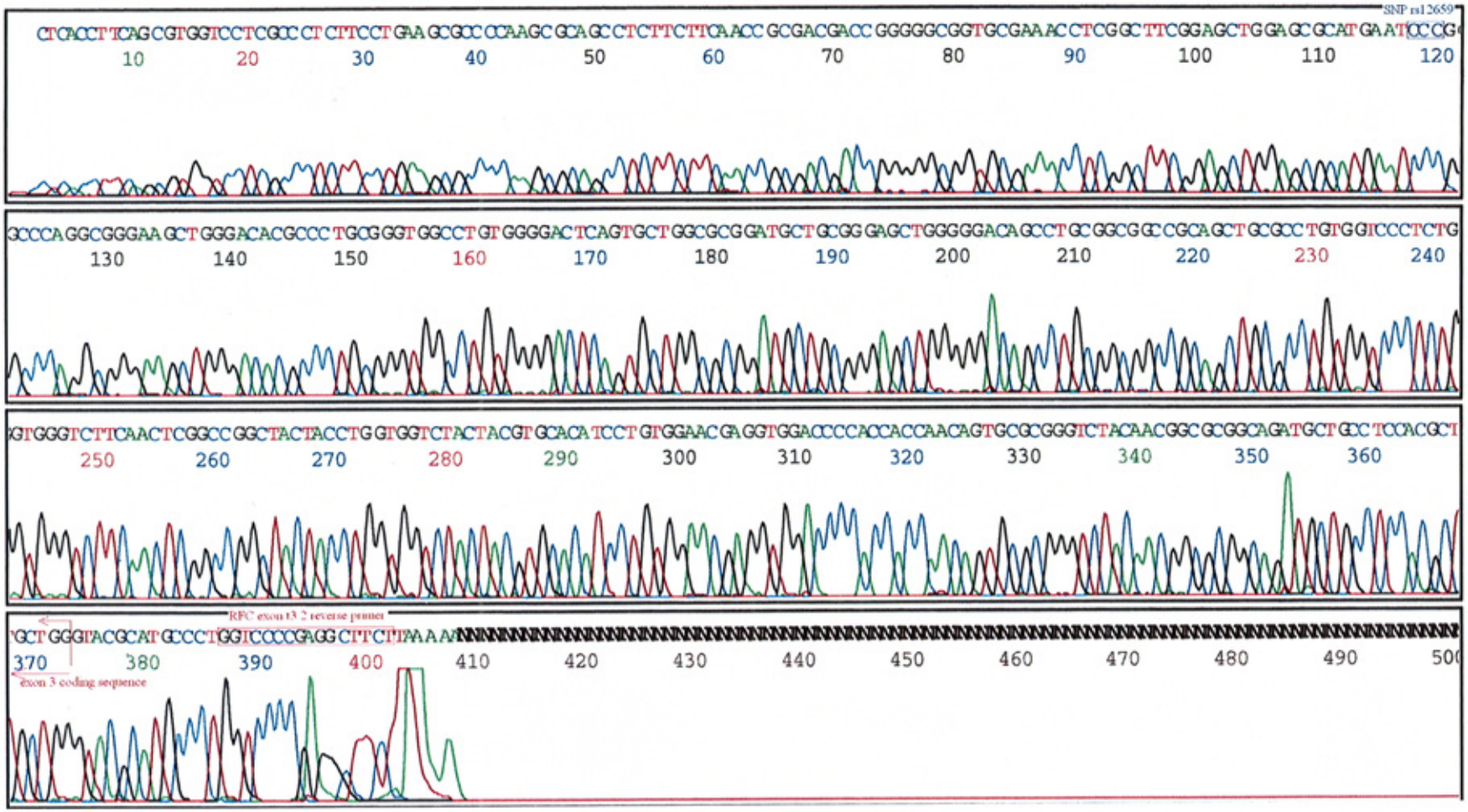
Appendix 3.2. Sequencing results from RFC exon 3.1 forward primer for NALM6 DNA.



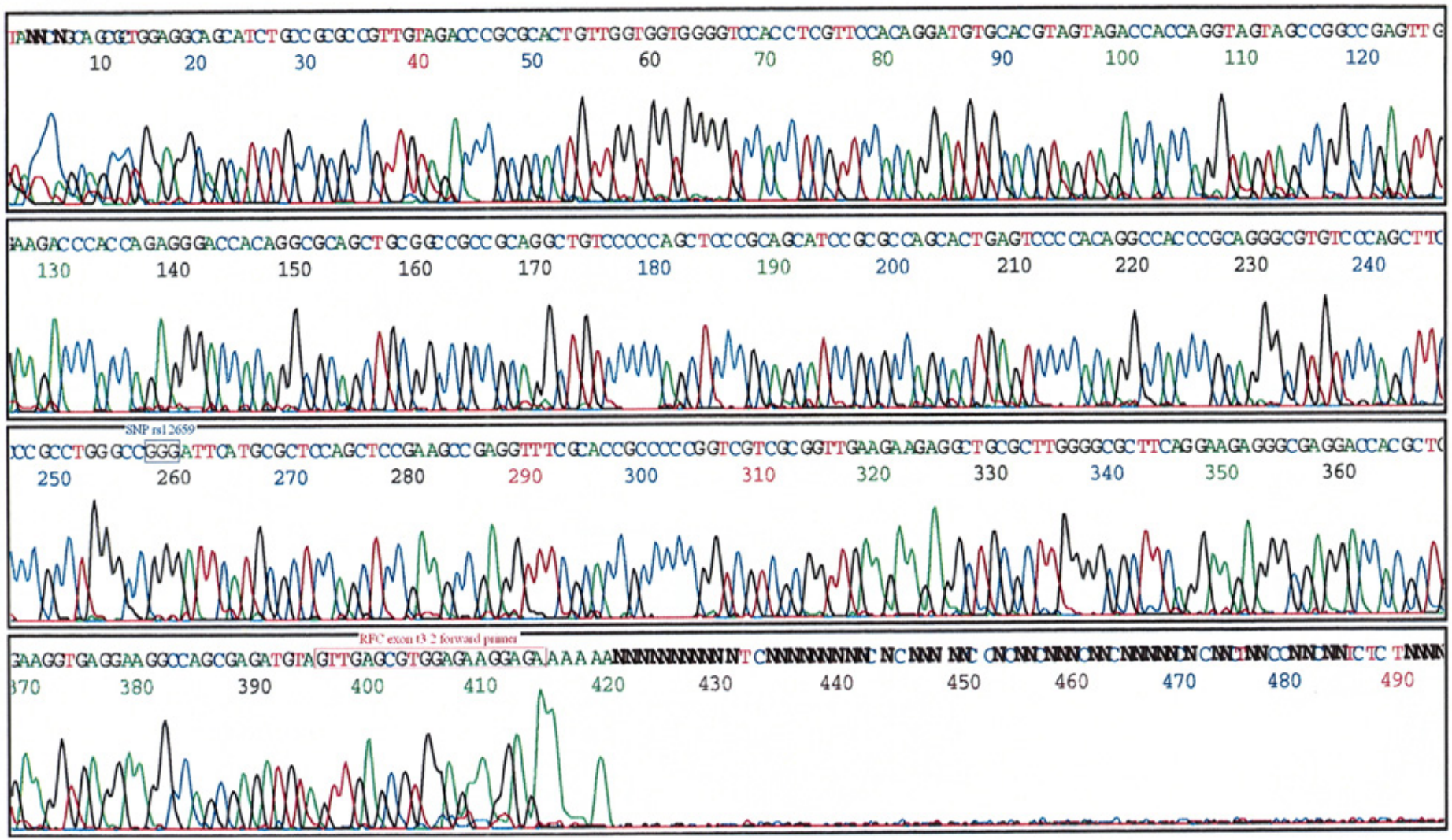
Appendix 3.3. Sequencing results from RFC exon 3.1 reverse primer for NALM6 DNA.



Appendix 3.4. Sequencing results from RFC exon t3.2 forward primer for NALM6 DNA.



Appendix 3.5. Sequencing results from RFC exon t3.2 reverse primer for NALM6 DNA.



Appendix 3.6 DHPLC results for RFC exon 3.1 at 65°C. Complete profiles are shown in green.

REDUCED FOLATE CARRIER EXON 3.1 **569bp**

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
76	homoduplex	homoduplex	WT	
171	homoduplex	homoduplex	WT	
279	homoduplex	homoduplex	WT	
L308	homoduplex	homoduplex	WT	
L345	homoduplex	homoduplex	WT	
L382	homoduplex	homoduplex	WT	
405	homoduplex	homoduplex	WT	
L609	heteroduplex			
980	homoduplex	homoduplex	WT	
1399	homoduplex	homoduplex	WT	
1533	homoduplex	homoduplex	WT	
1566	homoduplex	homoduplex	WT	
1720	homoduplex	homoduplex	WT	
1775	homoduplex	fail		
1785	homoduplex	homoduplex	WT	
2045	homoduplex	homoduplex	WT	
2416	homoduplex	homoduplex	WT	
2426	homoduplex	homoduplex	WT	
2578	homoduplex	homoduplex	WT	
2650	homoduplex	homoduplex	WT	
2865	homoduplex	homoduplex	WT	
3192	homoduplex	homoduplex	WT	
3451	homoduplex	homoduplex	WT	
3605	homoduplex	homoduplex	WT	
4225	homoduplex	homoduplex	WT	
4684	homoduplex	homoduplex	WT	

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
4845	homoduplex	homoduplex	WT	
4900	homoduplex	homoduplex	WT	
5467	homoduplex	homoduplex	WT	
5836	homoduplex	homoduplex	WT	
5862(31/7)	homoduplex	homoduplex	WT	
5862(27/9)	homoduplex	homoduplex	WT	
5887	homoduplex	homoduplex	WT	
6250	homoduplex	homoduplex	WT	
6346	homoduplex	homoduplex	WT	
6534	homoduplex	homoduplex	WT	
6563	homoduplex	homoduplex	WT	
6652	homoduplex	homoduplex	WT	
7096	homoduplex	homoduplex	WT	
8646	homoduplex	homoduplex	WT	
8668	homoduplex	homoduplex	WT	
8678	homoduplex	homoduplex	WT	
8679	homoduplex	homoduplex	WT	
8839	homoduplex	homoduplex	WT	
9216	homoduplex	homoduplex	WT	
9506	homoduplex	homoduplex	WT	

Continued

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
N1	homoduplex	homoduplex	WT	
N5	homoduplex	homoduplex	WT	
N7	homoduplex	homoduplex	WT	
N8	homoduplex	homoduplex	WT	
N11	homoduplex	homoduplex	WT	
N13	homoduplex	homoduplex	WT	
N14	homoduplex	homoduplex	WT	
N16	homoduplex	homoduplex	WT	
N17	homoduplex	homoduplex	WT	
N18	homoduplex	homoduplex	WT	
N20	homoduplex	homoduplex	WT	
N21	homoduplex	homoduplex	WT	
N22	homoduplex	homoduplex	WT	
N23	homoduplex	homoduplex	WT	
N24	homoduplex	homoduplex	WT	
N26	homoduplex	homoduplex	WT	
N27	homoduplex	homoduplex	WT	
N28	homoduplex	homoduplex	WT	
N29	homoduplex	homoduplex	WT	
N30	homoduplex	homoduplex	WT	
N31	homoduplex	homoduplex	WT	
N32	homoduplex	homoduplex	WT	
N36	homoduplex	homoduplex	WT	
N37	homoduplex	homoduplex	WT	
N38	homoduplex	homoduplex	WT	
N39	homoduplex	homoduplex	WT	
N40	homoduplex	homoduplex	WT	
N41	homoduplex	homoduplex	WT	
N45	homoduplex	homoduplex	WT	
N49	homoduplex	homoduplex	WT	

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
N50	homoduplex	homoduplex	WT	
N51	homoduplex	homoduplex	WT	
N52	homoduplex	homoduplex	WT	
N54	homoduplex	homoduplex	WT	
N56	homoduplex	homoduplex	WT	
N57	homoduplex	homoduplex	WT	
N58	homoduplex	homoduplex	WT	
N59	homoduplex	homoduplex	WT	
N60	homoduplex	homoduplex	WT	
N63	homoduplex	homoduplex	WT	
N64	homoduplex	homoduplex	WT	
N65	homoduplex	homoduplex	WT	
N68	homoduplex	homoduplex	WT	
N69	homoduplex	homoduplex	WT	
N70	homoduplex	homoduplex	WT	
N71	homoduplex	homoduplex	WT	
N72	homoduplex	homoduplex	WT	
N76	homoduplex	homoduplex	WT	
N77	heteroduplex			
N78	homoduplex	homoduplex	WT	
N79	homoduplex	homoduplex	WT	
N80	homoduplex	homoduplex	WT	
N81	homoduplex	homoduplex	WT	
N83	homoduplex	homoduplex	WT	

Appendix 3.7 DHPLC results for RFC exon t3.2 at 64.4, 65.4 and 66.4°C. Complete profiles at all three temperatures shown in green, samples which have enough data for genotyping of G80A polymorphism shown in yellow.

REDUCED FOLATE CARRIER Exon t3.2 441bp

	DHPLC at 64.4°C			DHPLC at 65.4°C			DHPLC at 66.4°C			
	Original	Spiked		Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
76	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
171	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
279	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
L308	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
L345	homoduplex	homoduplex	C/C	heteroduplex			fail			
L382	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
405	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
L609	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
980	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
1399	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
1533	heteroduplex		C/T	heteroduplex		C/T	heteroduplex		C/T	17/01/2006
1566	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
1720	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C				
1775	heteroduplex		C/T	heteroduplex			heteroduplex		C/T	26/08/2004
1785	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail	fail		
2416	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
2426	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
2578	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
2650	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
2865	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
3192	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
3451	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C				
3605	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
3990	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C				
4225	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C				
4684	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	

Continued

	DHPLC at 64.4°C			DHPLC at 65.4°C			DHPLC at 66.4°C			
	Original	**Spiked**		**Original**	**Spiked**		**Original**	**Spiked**		
Sample #	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**Date sequenced**
4900	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
5467	homoduplex	?heteroduplex	T/T	homoduplex	heteroduplex	T/T	homoduplex	?heteroduplex	T/T	
5836	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
5887	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail	fail		
6250	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
6652	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
7096	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C				
8646	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail	fail		
8668	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
8678	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
8679	homoduplex	homoduplex		homoduplex	fail		homoduplex	homoduplex	C/C	
8743	homoduplex	homoduplex	?	homoduplex	fail		homoduplex	homoduplex	C/C	26/08/2004
8839	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
8871	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C				
9212	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
9506	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
N5	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N8	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
N11	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N15	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	?homo		
N16	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
N17	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N18	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N19	fail			fail			homoduplex	homoduplex	C/C	
N20	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N21	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N23	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	

Continued

	DHPLC at 64.4°C			DHPLC at 65.4°C			DHPLC at 66.4°C			
	Original	**Spiked**		**Original**	**Spiked**		**Original**	**Spiked**		
Sample #	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**Date sequenced**
N24	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N26	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N27	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
N28	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N29	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N30	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N31	heteroduplex	?heteroduplex	C/T	fail			heteroduplex		C/T	
N32	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N36	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N38	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	fail			
N39	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N40	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N49	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
N50	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N51	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N52	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N57	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N60	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N63	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	26/08/2004
N64	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N65	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N69	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N70	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N71	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	26/08/2004
N72	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N73	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N74	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N76	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	

Continued

	DHPLC at 64.4°C			DHPLC at 65.4°C			DHPLC at 66.4°C			
	Original	**Spiked**		**Original**	**Spiked**		**Original**	**Spiked**		
Sample #	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**Date sequenced**
N77	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N78	homoduplex	homoduplex	C/C	homoduplex	fail		homoduplex	homoduplex	C/C	
N79	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	homoduplex	homoduplex	C/C	
N80	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N81	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	
N83	homoduplex	homoduplex	C/C	fail			homoduplex	homoduplex	C/C	

Appendix 3.8 χ^2-test to determine if the rs12659 genotype (C696T) distribution is significantly different between the relapse and normal groups.

	Observed				Expected			
Genotype	C/C	C/T	T/T	Total	C/C	C/T	T/T	Total
Normal	44	1	0	45	42.9	1.6	0.5	45
Relapse	37	2	1	40	38.1	1.4	0.5	40
Total	81	3	1	85	81	3	1	85

$\chi^2 = 0.458$. At 2 degrees of freedom this value is $0.10>P>0.05$ and therefore hypothesis (null hypothesis of no difference between observed and expected) cannot be rejected at 5% confidence.

Therefore there is no significant difference between distribution of genotypes between the normal and relapse data.

Appendix 4.1. Part of chromosome 21 amplified by RFC exon 4 primers (showing primers, coding sequence and key features).
Key:- Exon 4 shown in capital letters. Primers shown as shaded blocks. Other key features are shown in red.
Numbers across the top show coding base number using the first base in the start codon of exon 2 as position 1.

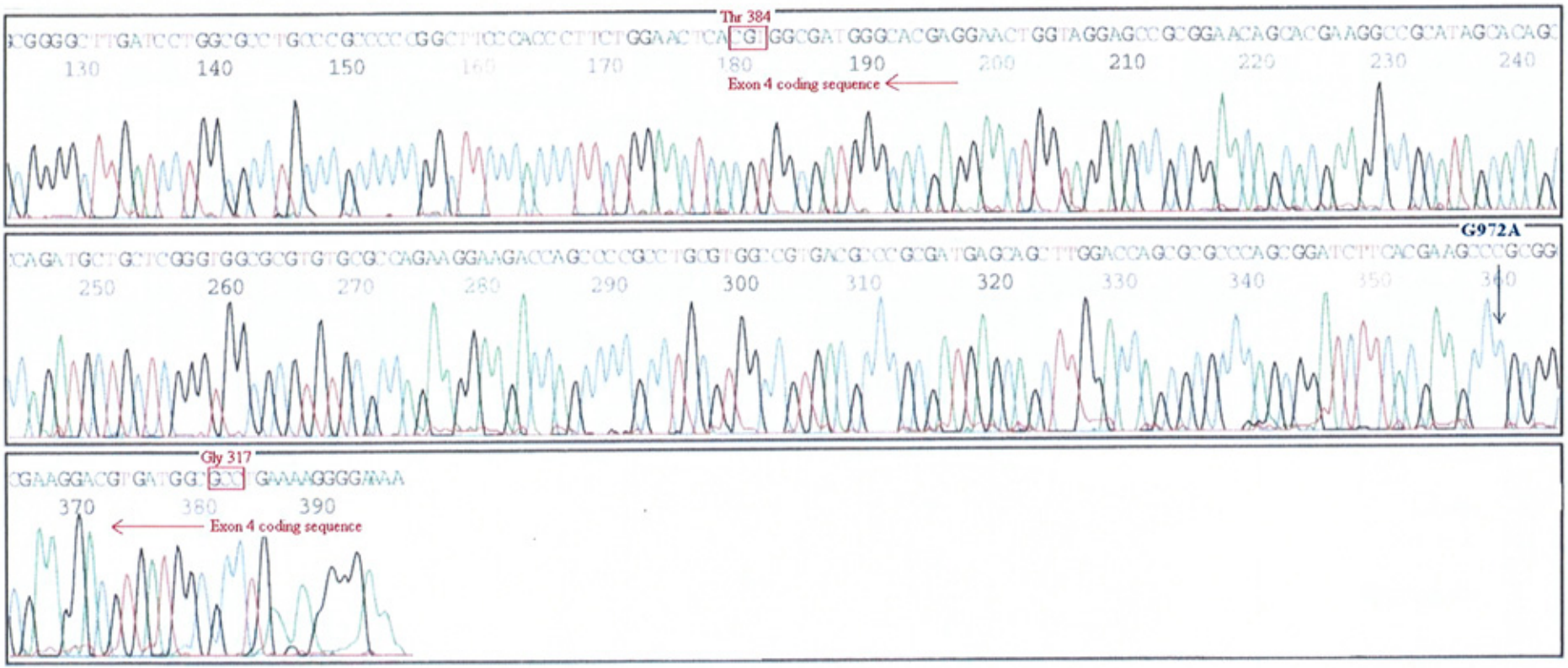
RFC exon 4 forward primer →

```
                949 951             961             971             981             991             1001            1011            1021            1031
c ccctctcaGG CGCCATCACG TCCTTCGCCG CGGGCTTCGT GAAGATCCGC TGGGCGCGCT GGTCCAAGCT GCTCATCGCG GGCGTCACGG CCACGCAGGC
                                    G972A SNP
                                proposed by Kaufmann et al
1041            1051            1061            1071            1081            1091            1101            1111            1121            1131
GGGGCTGGTC TTCCTTCTGG CGCACACGCG CCACCCGAGC AGCATCTGGC TGTGCTATGC GGCCTTCGTG CTGTTCCGCG GCTCCTACCA GTTCCTCGTG

1141            1152
CCCATCGCCA CGtgagttcc agaagggtgg gaagccgggg gcgggcaggc gccaggatca agccccgcgc ttcgctgggt gctcccacct tcctgctgcg ccccagcggc tgtgccactg

                                                                                ← RFC exon 4 reverse primer
ctggccccgc cagccacctg gagctgacca gggtccccac ttgcctctcc ctgggacctc gggcgtccct tcctgtgccg tcttgctgct cccccaggcc ctccac
```

Appendix 4.2. Sequencing results from RFC exon 4 forward primer for NALM6 DNA.

Sequencing starts well into the coding sequence, probably because the forward primer lies so close to the start of the exon and has interference.

Appendix 4.3. Sequencing results from RFC exon 4 reverse primer for NALM6 DNA. Position of Kaufmann *et al* proposed G972A SNP shown in blue, first and last residues coded by exon 4 are shown in red.



Appendix 4.4 DHPLC results for RFC exon 4 at 65.9°C. Complete profiles are shown in green.

REDUCED FOLATE CARRIER EXON 4 **427 bp**

	Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	Date sequenced
76	homoduplex	homoduplex	G/G	
171	homoduplex	homoduplex	G/G	
279	homoduplex	homoduplex	G/G	
L308	homoduplex	homoduplex	G/G	
L345	homoduplex	homoduplex	G/G	
L382	homoduplex	homoduplex	G/G	
405	homoduplex	homoduplex	G/G	
L609	heteroduplex		? G/A	
980	homoduplex	homoduplex	G/G	
981	homoduplex	homoduplex	G/G	
1399	homoduplex	homoduplex	G/G	
1533	homoduplex	homoduplex	G/G	
1566	homoduplex	homoduplex	G/G	
1720	homoduplex	homoduplex	G/G	
1775	homoduplex	homoduplex	G/G	
1785	homoduplex	homoduplex	G/G	
2416	homoduplex	homoduplex	G/G	
2426	homoduplex	homoduplex	G/G	
2578	homoduplex	homoduplex	G/G	
2650	homoduplex	homoduplex	G/G	
2865	homoduplex	homoduplex	G/G	
3192	homoduplex	homoduplex	G/G	
3451	homoduplex	homoduplex	G/G	
3605	homoduplex	homoduplex	G/G	25/11/2003
3990	homoduplex	homoduplex	G/G	
4225	homoduplex	homoduplex	G/G	

	Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	Date sequenced
4684	homoduplex	homoduplex	G/G	
4900	homoduplex	homoduplex	G/G	
5467	homoduplex	homoduplex	G/G	
5836	homoduplex	homoduplex	G/G	
5862(27/9)	homoduplex	homoduplex	G/G	
6250	homoduplex	homoduplex	G/G	
6346	homoduplex	homoduplex	G/G	
6563	homoduplex	homoduplex	G/G	
6652	homoduplex	homoduplex	G/G	
7096	homoduplex	homoduplex	G/G	
8646	homoduplex	homoduplex	G/G	
8668	homoduplex	homoduplex	G/G	
8678	homoduplex	homoduplex	G/G	
8679	homoduplex	homoduplex	G/G	
8839	homoduplex	homoduplex	G/G	
8871	homoduplex	fail		
9216	homoduplex	homoduplex	G/G	
9506	homoduplex	homoduplex	G/G	
11420	homoduplex	homoduplex	G/G	

Continued

Continued

	Original	Spiked		
Sample #	**WAVE result**	**WAVE result**	**Genotype**	**Date sequenced**
N1	homoduplex	homoduplex	G/G	
N3	homoduplex	homoduplex	G/G	
N5	homoduplex	homoduplex	G/G	
N7	homoduplex	homoduplex	G/G	
N8	homoduplex	homoduplex	G/G	
N12	homoduplex	heteroduplex	?A/A	
N15	homoduplex	homoduplex	G/G	
N17	homoduplex	homoduplex	G/G	
N18	homoduplex	homoduplex	G/G	
N20	homoduplex	homoduplex	G/G	
N21	homoduplex	homoduplex	G/G	
N22	homoduplex	homoduplex	G/G	
N23	homoduplex	homoduplex	G/G	
N24	homoduplex	homoduplex	G/G	
N25	homoduplex	homoduplex	G/G	25/11/2003
N26	homoduplex	homoduplex	G/G	
N27	homoduplex	homoduplex	G/G	
N28	homoduplex	homoduplex	G/G	
N29	homoduplex	homoduplex	G/G	
N30	homoduplex	homoduplex	G/G	
N31	homoduplex	homoduplex	G/G	
N32	homoduplex	homoduplex	G/G	
N33	heteroduplex		??	
N34	homoduplex	homoduplex	G/G	
N35	homoduplex	homoduplex	G/G	
N36	heteroduplex		G/A	19/11/2004
N37	homoduplex	homoduplex	G/G	
N38	homoduplex	homoduplex	G/G	
N39	homoduplex	homoduplex	G/G	

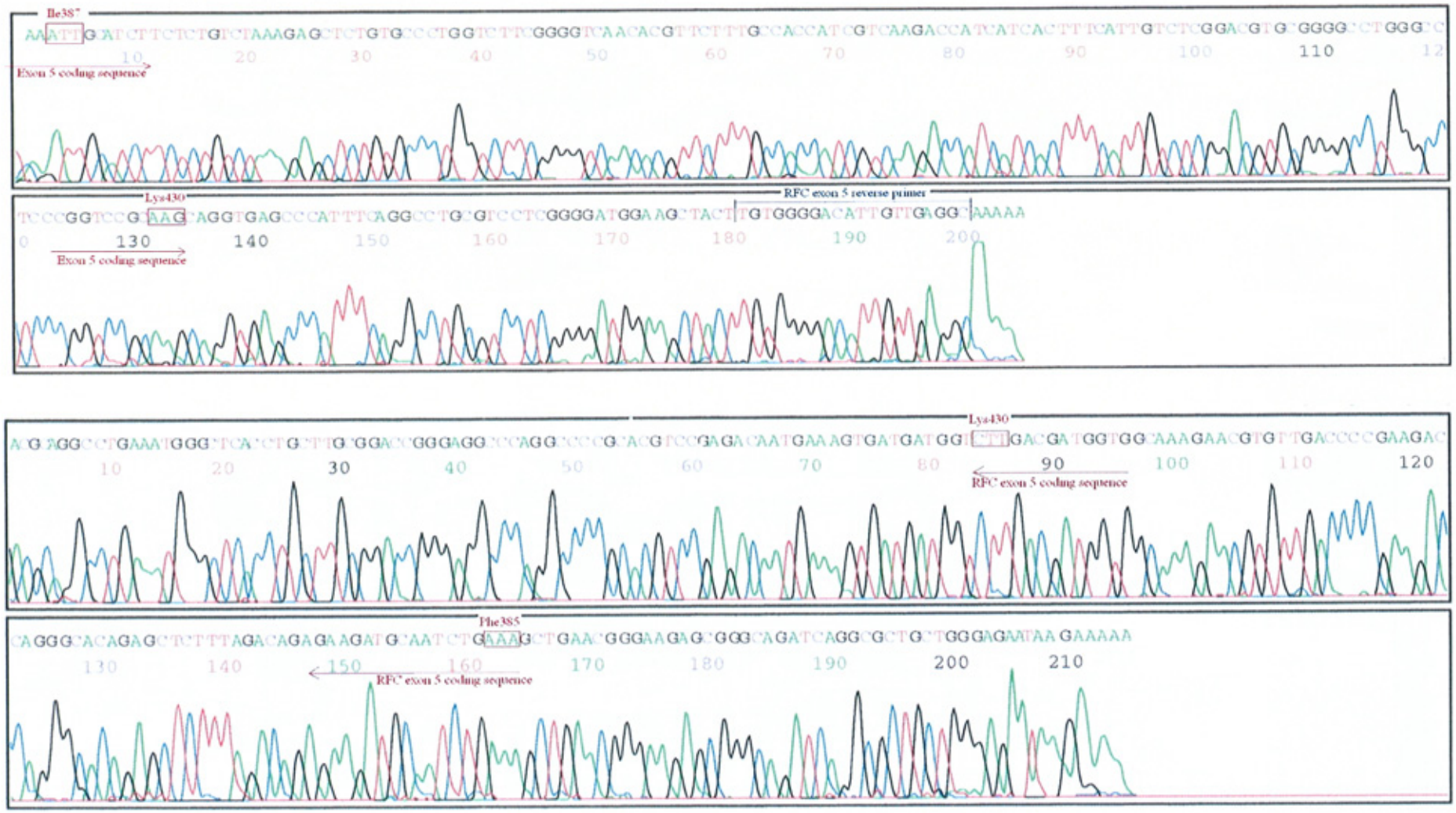
	Original	Spiked		
Sample #	**WAVE result**	**WAVE result**	**Genotype**	**Date sequenced**
N40	homoduplex	homoduplex	G/G	25/11/2003
N49	homoduplex	homoduplex	G/G	
N50	homoduplex	homoduplex	G/G	
N51	homoduplex	homoduplex	G/G	
N52	homoduplex	homoduplex	G/G	
N57	homoduplex	homoduplex	G/G	
N60	homoduplex	homoduplex	G/G	
N63	homoduplex	homoduplex	G/G	
N64	homoduplex	homoduplex	G/G	
N65	homoduplex	homoduplex	G/G	
N68	homoduplex	homoduplex	G/G	
N69	homoduplex	homoduplex	G/G	
N70	homoduplex	homoduplex	G/G	
N71	homoduplex	homoduplex	G/G	
N72	homoduplex	homoduplex	G/G	
N74	homoduplex	homoduplex	G/G	
N76	homoduplex	homoduplex	G/G	
N77	homoduplex	homoduplex	G/G	
N78	homoduplex	homoduplex	G/G	
N79	homoduplex	homoduplex	G/G	
N80	homoduplex	homoduplex	G/G	
N81	homoduplex	homoduplex	G/G	
N83	homoduplex	homoduplex	G/G	

Appendix 5.1. Part of chromosome 21 amplified by RFC exon 5 primers (showing primers, coding sequence and key features).
Key:- Exon 5 shown in capital letters. Primers shown as shaded blocks.
Numbers across the top show coding base number using the first base in the start codon of exon 2 as position 1.

RFC exon 5 forward primer →

```
                                1153            1161            1171            1181            1191
ctta ttctcccagc agcgcctgat ctgcccgctc ttcccgttca GCTTTCAGAT TGCATCTTCT CTGTCTAAAG AGCTCTGTGC CCTGGTCTTC

1201            1211            1221            1231            1241            1251            1261            1271
GGGGTCAACA CGTTCTTTGC CACCATCGTC AAGACCATCA TCACTTTCAT TGTCTCGGAC GTGCGGGGCC TGGGCCTCCC

1281        1290                                                ← RFC exon 5 forward primer
GGTCCGCAAG caggtgagcc catttcaggc ctgcgtcctc ggggatggaa gctacttgtg gggacattgt tgaggc
```

Appendix 5.2. Sequencing results from RFC exon 5 forward primer (top) and reverse primer (bottom) for NALM6 DNA. Key features are shown.



Appendix 5.3 DHPLC results for RFC exon 5 at 62°C. Complete profiles shown in green.

REDUCED FOLATE CARRIER EXON 5 **250 bp**

	Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	Date sequenced
L68	homoduplex	homoduplex	W/T	
76	homoduplex	homoduplex	W/T	
171	homoduplex	homoduplex	W/T	
279	homoduplex	homoduplex	W/T	
405	homoduplex	homoduplex	W/T	
980	homoduplex	homoduplex	W/T	
981	homoduplex	homoduplex	W/T	
1399	homoduplex	homoduplex	W/T	
1533	homoduplex	homoduplex	W/T	
1566	homoduplex	homoduplex	W/T	
1720	homoduplex	homoduplex	W/T	
1775	homoduplex	homoduplex	W/T	
1785	homoduplex	homoduplex	W/T	
2045	homoduplex	homoduplex	W/T	
2416	homoduplex	homoduplex	W/T	
2426	homoduplex	homoduplex	W/T	
2578	homoduplex	homoduplex	W/T	
2650	homoduplex	homoduplex	W/T	
2865	homoduplex	homoduplex	W/T	
3192	homoduplex	homoduplex	W/T	
3451	homoduplex	homoduplex	W/T	
3605	homoduplex	homoduplex	W/T	
3990	homoduplex	homoduplex	W/T	
4225	homoduplex	homoduplex	W/T	

	Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	Date sequenced
4684	homoduplex	homoduplex	W/T	
4845	homoduplex	homoduplex	W/T	
4900	homoduplex	homoduplex	W/T	
5467	homoduplex	homoduplex	W/T	
5836	homoduplex	homoduplex	W/T	
5862(31/7)	homoduplex	homoduplex	W/T	
5862(27/9)	homoduplex	homoduplex	W/T	
5887	homoduplex	homoduplex	W/T	
6250	homoduplex	homoduplex	W/T	
6563	homoduplex	homoduplex	W/T	
6652	homoduplex	homoduplex	W/T	
7096	homoduplex	homoduplex	W/T	
8646	homoduplex	homoduplex	W/T	
8668	homoduplex	homoduplex	W/T	
8678	homoduplex	homoduplex	W/T	
8679	homoduplex	homoduplex	W/T	
8839	homoduplex	homoduplex	W/T	
8871	homoduplex	homoduplex	W/T	
9216	homoduplex	homoduplex	W/T	
9506	homoduplex	homoduplex	W/T	
11420	homoduplex	homoduplex	W/T	

Continued

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
N1	homoduplex	homoduplex	W/T	
N2	homoduplex	homoduplex	W/T	
N3	homoduplex	homoduplex	W/T	
N4	homoduplex	homoduplex	W/T	
N5	homoduplex	homoduplex	W/T	
N6	homoduplex	homoduplex	W/T	
N8	homoduplex	homoduplex	W/T	
N9	homoduplex	homoduplex	W/T	
N10	homoduplex	homoduplex	W/T	
N11	homoduplex	homoduplex	W/T	
N12	homoduplex	homoduplex	W/T	
N13	homoduplex	homoduplex	W/T	
N14	homoduplex	homoduplex	W/T	
N15	homoduplex	homoduplex	W/T	
N16	homoduplex	homoduplex	W/T	
N17	homoduplex	homoduplex	W/T	
N18	homoduplex	homoduplex	W/T	
N19	homoduplex	homoduplex	W/T	
N21	homoduplex	homoduplex	W/T	
N22	homoduplex	homoduplex	W/T	
N23	homoduplex	homoduplex	W/T	
N24	homoduplex	homoduplex	W/T	
N25	homoduplex	homoduplex	W/T	
N26	homoduplex	homoduplex	W/T	
N27	homoduplex	homoduplex	W/T	
N28	homoduplex	homoduplex	W/T	
N29	homoduplex	homoduplex	W/T	
N30	homoduplex	homoduplex	W/T	
N32	homoduplex	homoduplex	W/T	
N33	homoduplex	homoduplex	W/T	

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
N34	homoduplex	homoduplex	W/T	
N35	homoduplex	homoduplex	W/T	
N36	homoduplex	homoduplex	W/T	
N37	homoduplex	homoduplex	W/T	
N38	homoduplex	homoduplex	W/T	
N39	homoduplex	homoduplex	W/T	
N40	homoduplex	homoduplex	W/T	
N49	homoduplex	homoduplex	W/T	
N50	homoduplex	homoduplex	W/T	
N51	homoduplex	homoduplex	W/T	
N52	homoduplex	homoduplex	W/T	
N57	homoduplex	homoduplex	W/T	
N60	homoduplex	homoduplex	W/T	
N63	homoduplex	homoduplex	W/T	
N64	homoduplex	homoduplex	W/T	
N65	homoduplex	homoduplex	W/T	
N68	homoduplex	homoduplex	W/T	
N69	homoduplex	homoduplex	W/T	
N70	homoduplex	homoduplex	W/T	
N71	homoduplex	homoduplex	W/T	
N72	homoduplex	homoduplex	W/T	
N73	homoduplex	homoduplex	W/T	
N74	homoduplex	homoduplex	W/T	
N76	homoduplex	homoduplex	W/T	
N77	homoduplex	homoduplex	W/T	
N78	homoduplex	homoduplex	W/T	
N79	homoduplex	homoduplex	W/T	
N80	homoduplex	homoduplex	W/T	
N81	homoduplex	homoduplex	W/T	
N83	homoduplex	homoduplex	W/T	

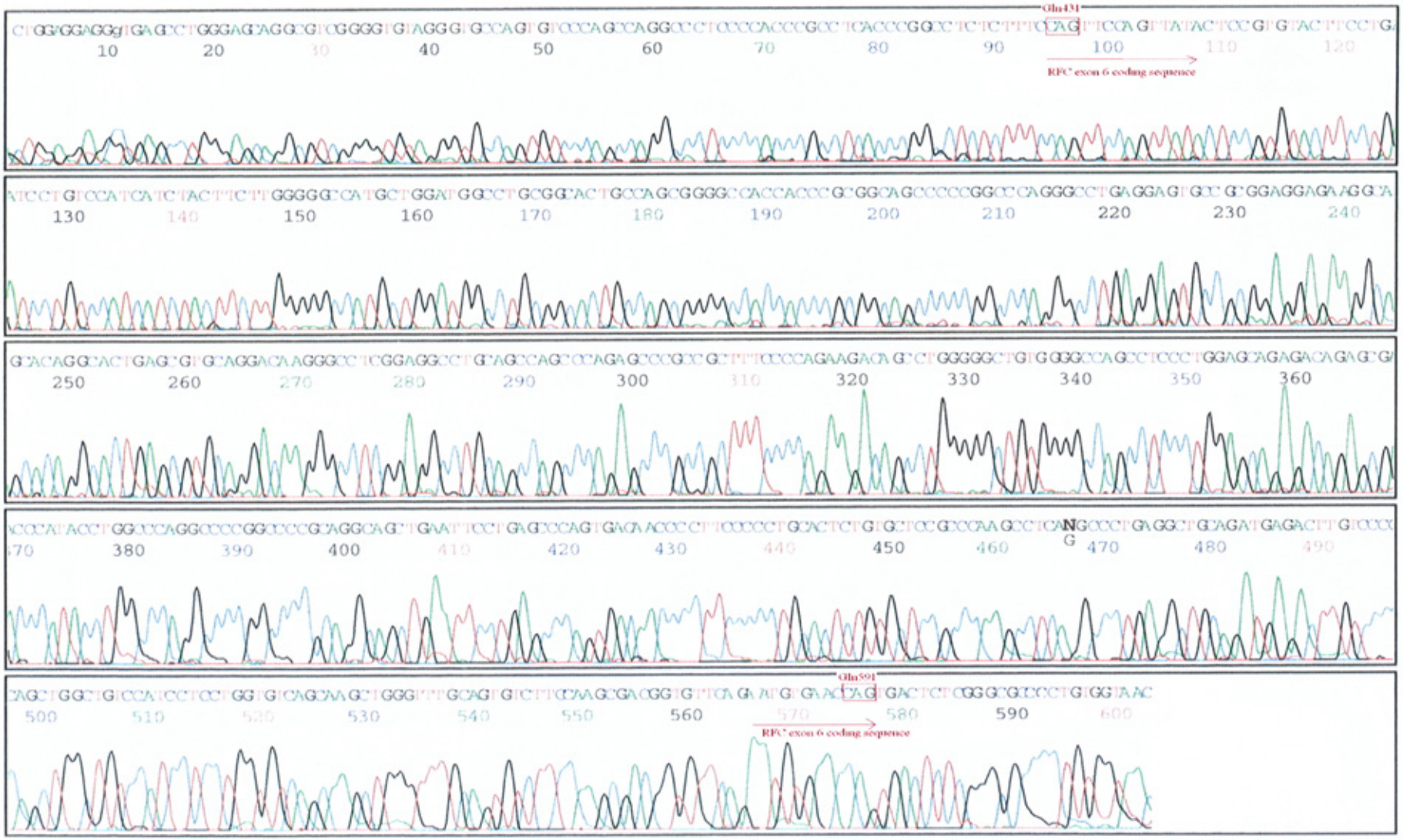
Appendix 6.1. Part of chromosome 21 amplified by RFC exon p6 primers (showing primers, coding sequence and key features).
Key:- Exon shown in capital letters. Primers shown as shaded blocks. Other key features are shown in red.
Numbers across the top show coding base number using the first base in the start codon of exon 2 as position 1.

RFC exon 6 forward primer →

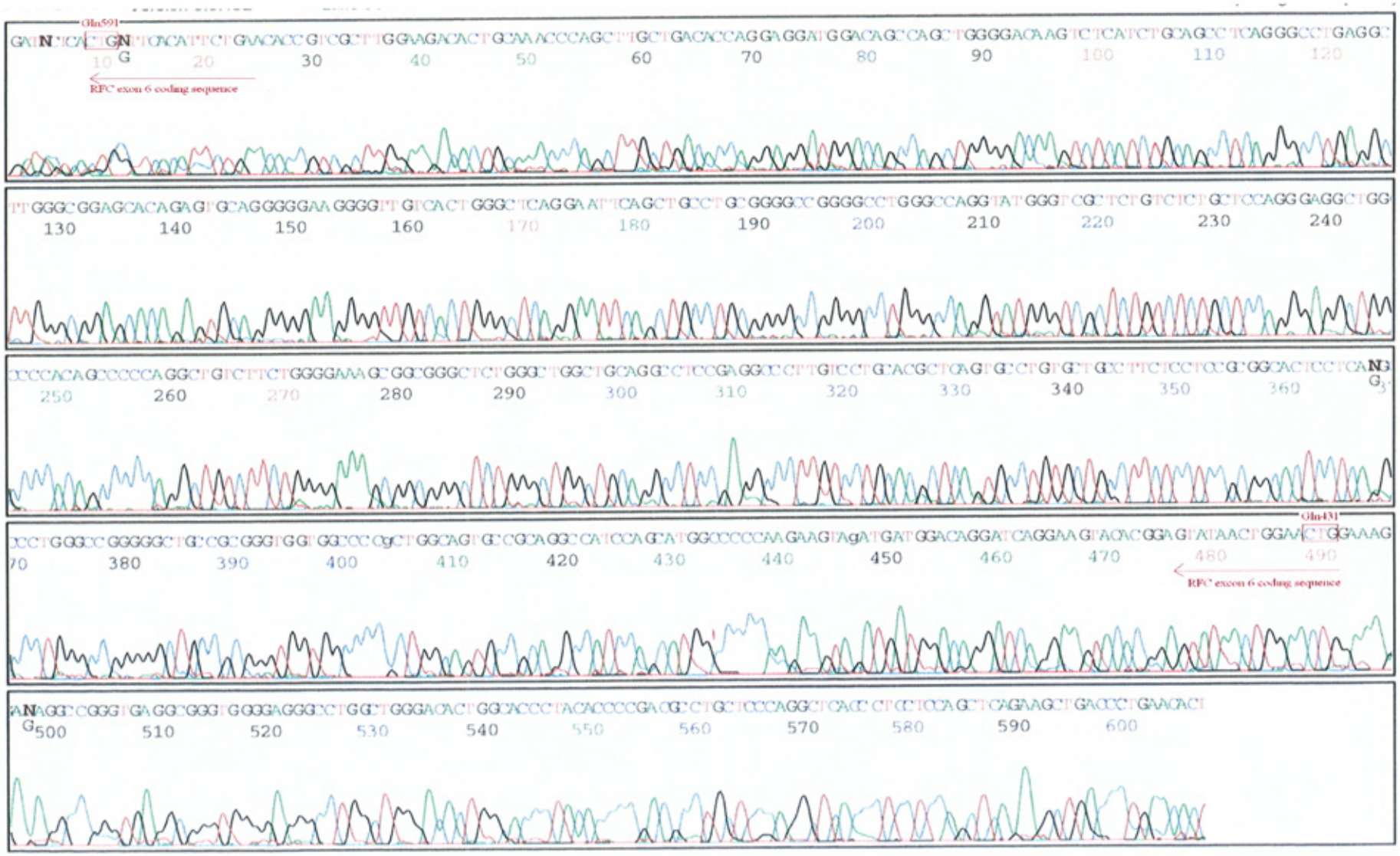
tgaccgctg tgcagtgttc agggtcagcc tctgagctgg aggagggtga gcctgggagc aggcgtcggg gtgtagggtg ccagtgtccc agccaggccc tccccacccg cctcacccgg

```
          1291                1301                1311                1321                1331                1341                1351
cctctctttc CAGTTCCAGT TATACTCCGT GTACTTCCTG ATCCTGTCCA TCATCTACTT CTTGGGGGCC ATGCTGGATG

1361                1371                1381                1391                1401                1411                1421                1431
GCCTGCGGCA CTGCCAGCGG GGCCACCACC CGCGGCAGCC CCCGGCCCAG GGCCTGAGGA GTGCCGCGGA GGAGAAGGCA
                                                    C1401T SNP rs7278825
1441                1451                1461                1471                1481                1491                1501                1511
GCACAGGCAC TGAGCGTGCA GGACAAGGGC CTCGGAGGCC TGCAGCCAGC CCAGAGCCCG CCGCTTTCCC CAGAAGACAG

1521                1531                1541                1551                1561                1571                1581                1591
CCTGGGGGCT GTGGGGCCAG CCTCCCTGGA GCAGAGACAG AGCGACCCAT ACCTGGCCCA GGCCCCGGCC CCGCAGGCAG

1601                1611                1621                1631                1641                1651                1661                1671
CTGAATTCCT GAGCCCAGTG ACAACCCCTT CCCCCTGCAC TCTGTGCTCC GCCCAAGCCT CAGGCCCTGA GGCTGCAGAT

1681                1691                1701                1711                1721                1731                1741                1751
GAGACTTGTC CCCAGCTGGC TGTCCATCCT CCTGGTGTCA GCAAGCTGGG TTTGCAGTGT CTTCCAAGCG ACGGTGTTCA

1761                1773                ← RFC exon 6 forward primer
GAATGTGAAC CAGtgactct cgggcgcccc tgtggtaact ttgcaggcg
```

Appendix 6.2. Sequencing results from RFC exon 6 forward primers for NALM6 DNA.



Appendix 6.3. Sequencing results from RFC exon 6 reverse primer for NALM6 DNA.



Appendix 6.4 DHPLC results for RFC exon 6 at 60.5, 63.0, 66.0 and 68.0°C. Complete profiles are shown in green.

REDUCED FOLATE CARRIER Exon p6 **649 bp**

	DHPLC at 60.5°C			DHPLC at 63.0°C			DHPLC at 66.0°C			DHPLC at 68.0°C			
	Original	**Spiked**		**Original**	**Spiked**		**Original**	**Spiked**		**Original**	**Spiked**		
Sample #	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**Date sequenced**
76	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
171	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
279	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
L308				homoduplex	homoduplex	W/T							
L345				homoduplex	homoduplex	W/T							
L382				homoduplex	homoduplex	W/T							
405	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
L609				homoduplex	homoduplex	W/T							
980				homoduplex	homoduplex	W/T							
1339	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
1533	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
1720	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
1775	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
1785	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
1920	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
2416	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
2426	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
2578	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
2650	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
2865	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
3016	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
3192	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
3451	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
3605	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
3990	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
4684	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	

Continued

	DHPLC at 60.5°C			DHPLC at 63.0°C			DHPLC at 66.0°C			DHPLC at 68.0°C			
	Original	Spiked		Original	Spiked		Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
4900	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
5467	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
5836	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
6250	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
6563	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
6652	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
6845	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
7096	homoduplex	heteroduplex	?	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	17/1/2006
8646	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
8668	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
8678	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
8679	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
8743	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
8839	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
9212	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
9216	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
9506	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N1	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N5	homoduplex	homoduplex	W/T	fail	homoduplex		fail	homoduplex			homoduplex		
N8	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N15	homoduplex	homoduplex	W/T	fail	homoduplex		fail	homoduplex			homoduplex		
N18	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N21	homoduplex	homoduplex	W/T	fail	homoduplex		fail	homoduplex			homoduplex		
N23	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N26	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	

REDUCED FOLATE CARRIER Exon p6 **649 bp**

	DHPLC at 60.5°C			DHPLC at 63.0°C			DHPLC at 66.0°C			DHPLC at 68.0°C			
	Original	**Spiked**		**Original**	**Spiked**		**Original**	**Spiked**		**Original**	**Spiked**		
Sample #	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**Date sequenced**
N27	homoduplex	homoduplex	W/T	fail	homoduplex		homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N28	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N29	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N30	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N32	homoduplex	homoduplex	W/T	fail	homoduplex		fail	homoduplex			homoduplex		
N36	homoduplex	heteroduplex	?	homoduplex	heteroduplex	?	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	17/1/2006
N38	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N39	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N40	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N49	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N50	homoduplex	heteroduplex	?	homoduplex	heteroduplex	?	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	17/1/2006
N52	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N57	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N60	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N63	homoduplex			heteroduplrx			heteroduplex	fail		heteroduplex			
N65	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N68	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N69	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N71	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N78	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N79	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N81	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	
N83	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	homoduplex	homoduplex	W/T	

Appendix 7.1. Part of chromosome 5 amplified by DHFR exon m2 primers (showing primers, coding sequence and key features).

Key:- Exon 2 shown in capital letters. Primers shown as shaded blocks. Other key features are shown in red.
Numbers across the top show base number using the first base in the start codon (shown blue) as position 1.

Sample L382
A/G -25

cgcctgcac aaatggggac gaggggggcg gggcggcca caatttcgcg ccaaacttga ccgcgcgttc tgctgtaacg agcgggctcg gaggtcctcc cgctgctgtc

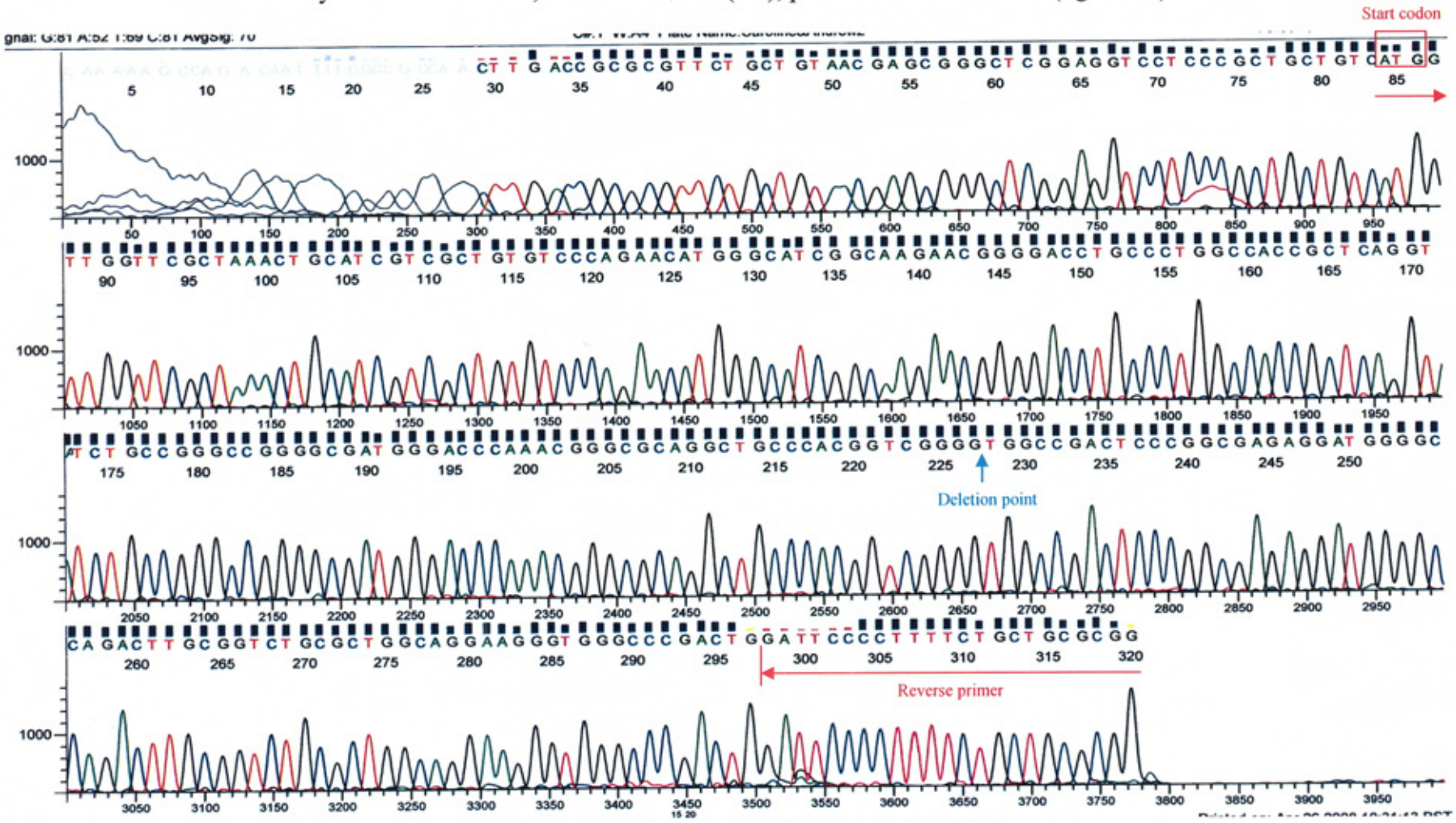
1 11 21 31 41 51 61 71
ATGGTTGGTT CGCTAAACTG CATCGTCGCT GTGTCCCAGA ACATGGGCAT CGGCAAGAAC GGGGACCTGC CCTGGCCACC
Start codon

81 87 ↓ position of 19 base pair deletion
GCTCAGGtat ctgccgggcc ggggcgatgg gacccaaacg ggcgcaggct gcccacggtcg gggtggccga ctcccggcgag aggatggggc cagacttgcgg tctgcgctggc

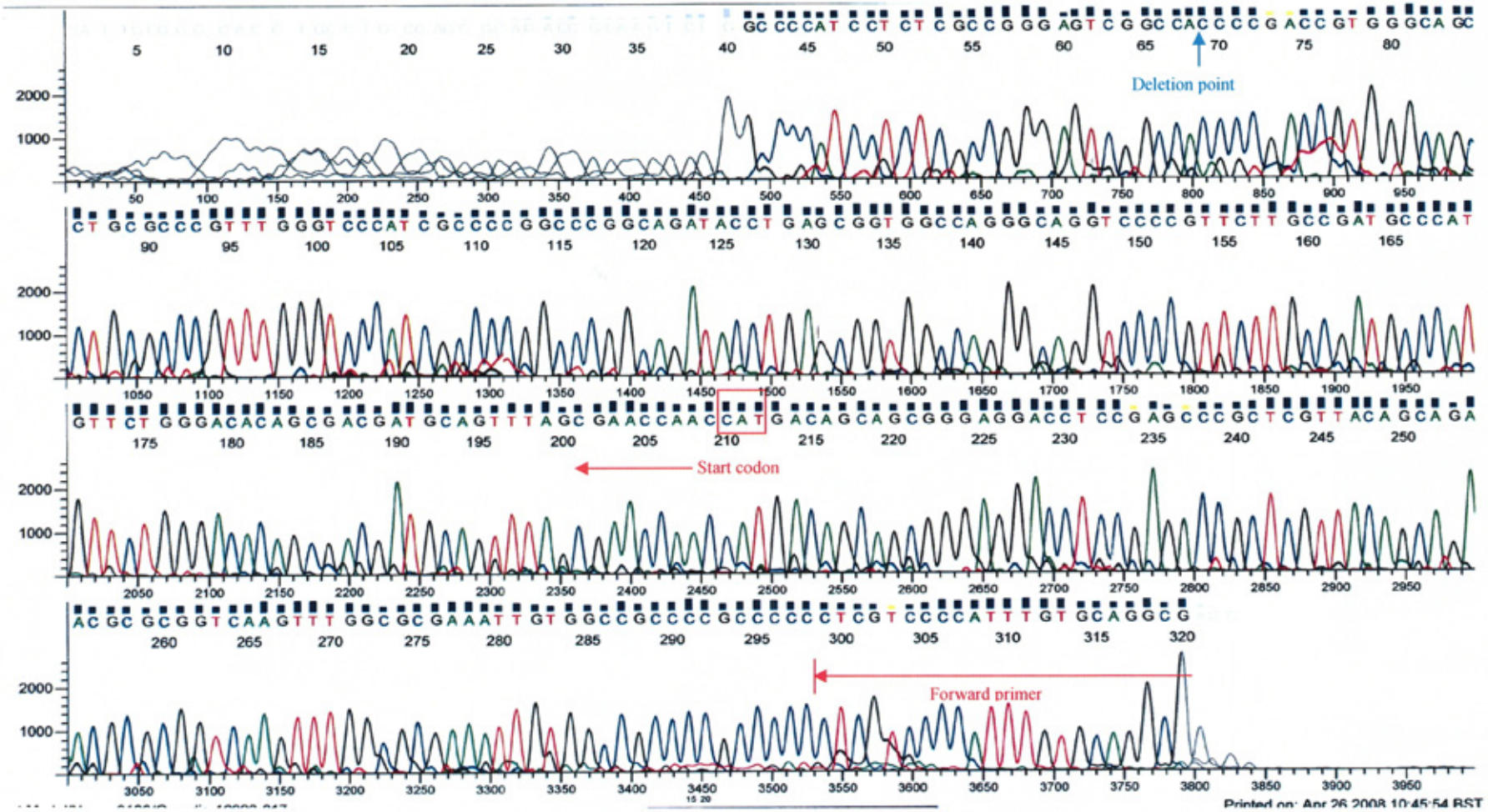
Sample 6625
A/G

aggaagggtgg gcccgactgg attccccttt tctgctgcgcg g

Appendix 7.2 Sequencing results from DHFR exon m2 forward primer (top) for CCRF-CEM DNA.
Key features are shown; ATG start codon (red); point of 19 base deletion (light blue).



Appendix 7.3 Sequencing results from DHFR exon m2 reverse primer for CCRF-CEM DNA.
Key features are shown; ATG start codon (red); position of 19 base deletion (light blue).



Appendix 7.4 DHPLC results for DHFR exon2 at 65°C. Complete profiles shown in green.

DIHYDROFOLATE REDUCTASE EXON 2 **345 bp**

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
74(L76)	Heteroduplex (3)		+19/-19	
360(L279)	Heteroduplex (3)		+19/-19	
189(L405)	Heteroduplex (3)		+19/-19	
L308	Heteroduplex (3)		+19/-19	
L345	Heteroduplex (3)		+19/-19	
L382	Heteroduplex (2)		?	26/4/08
L609	Heteroduplex (2)			
980(L213)	Heteroduplex (3)		+19/-19	
981(L213)	Heteroduplex (3)		+19/-19	
1533	Heteroduplex (4)		?	26/4/08
1775	Heteroduplex (3)		+19/-19	
1785	Heteroduplex (3)		+19/-19	
2416	Heteroduplex (3)		+19/-19	
2426	Heteroduplex (3)		+19/-19	
2650	Heteroduplex (3)		+19/-19	
2865	Heteroduplex (2)			
3192	Heteroduplex (3)		+19/-19	
3451	Heteroduplex (3)		+19/-19	
3605	? Heteroduplex			
4684	Heteroduplex (3)		+19/-19	
5467	Heteroduplex (3)		+19/-19	
5862(31/7)	Heteroduplex (3)		+19/-19	
6250	Homoduplex	Homoduplex	-19/-19	
6563	Heteroduplex (3)		+19/-19	
6652	Heteroduplex (2)		?	26/4/08

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
7096	Heteroduplex (3)		+19/-19	
8678	Heteroduplex (3)		+19/-19	
8839	Homoduplex	Homoduplex	-19/-19	
9506	Heteroduplex (3)		+19/-19	
11420	Heteroduplex (3)		+19/-19	
N4	Homoduplex	Heteroduplex	+19/+19	
N8	Homoduplex	Heteroduplex	+19/+19	
N11	Heteroduplex (3)		+19/-19	
N12	Homoduplex	Fail		
N13	Homoduplex	Heteroduplex	+19/+19	
N18	Heteroduplex (3)		+19/-19	
N28	Heteroduplex (3)		+19/-19	
N38	Homoduplex	Fail		
N50	Heteroduplex (3)		+19/-19	
N51	Homoduplex	Fail		
N52	Heteroduplex (3)		+19/-19	
N56	Homoduplex	Homoduplex	-19/-19	
N65	Homoduplex	Heteroduplex	+19/+19	
N72	Homoduplex	Homoduplex	-19/-19	
N74	Heteroduplex (2)			
N77	Heteroduplex (3)		+19/-19	
N81	Homoduplex	Homoduplex	-19/-19	
N83	Heteroduplex (3)		+19/-19	
N84	Heteroduplex (3)		+19/-19	

Appendix 7.5 χ^2-test to determine if the genotype distribution is significantly different between the relapse and normal groups. Genotypes grouped according to Johnson *et al* (2005)

	Observed			Expected		
Genotype	WT/WT	WT/-19 or -19/-19	Total	WT/WT	WT/-19 or -19/-19	Total
Normal	0	24	24	2.46	21.54	24
Relapse	4	11	15	1.54	13.46	15
Total	4	35	39	4.00	35.00	39

$\chi^2 = 9.765$. At 1 degree of freedom this value has a P value of 0.0076 and therefore the null hypothesis (no difference) can be rejected. Because of the small sample size the statistical analysis is not strong evidence, but there does appear to be a difference between the normal and relapse groups that requires further investigation.

Appendix 8.1. Part of chromosome 5 amplified by DHFR exon 3 primers (showing primers, coding sequence and key features).

Key:- Exon 3 shown in capital letters. Primers shown as shaded blocks. Other key features are shown in red.
Numbers across the top show coding base number using the first base in the start codon as position 1.

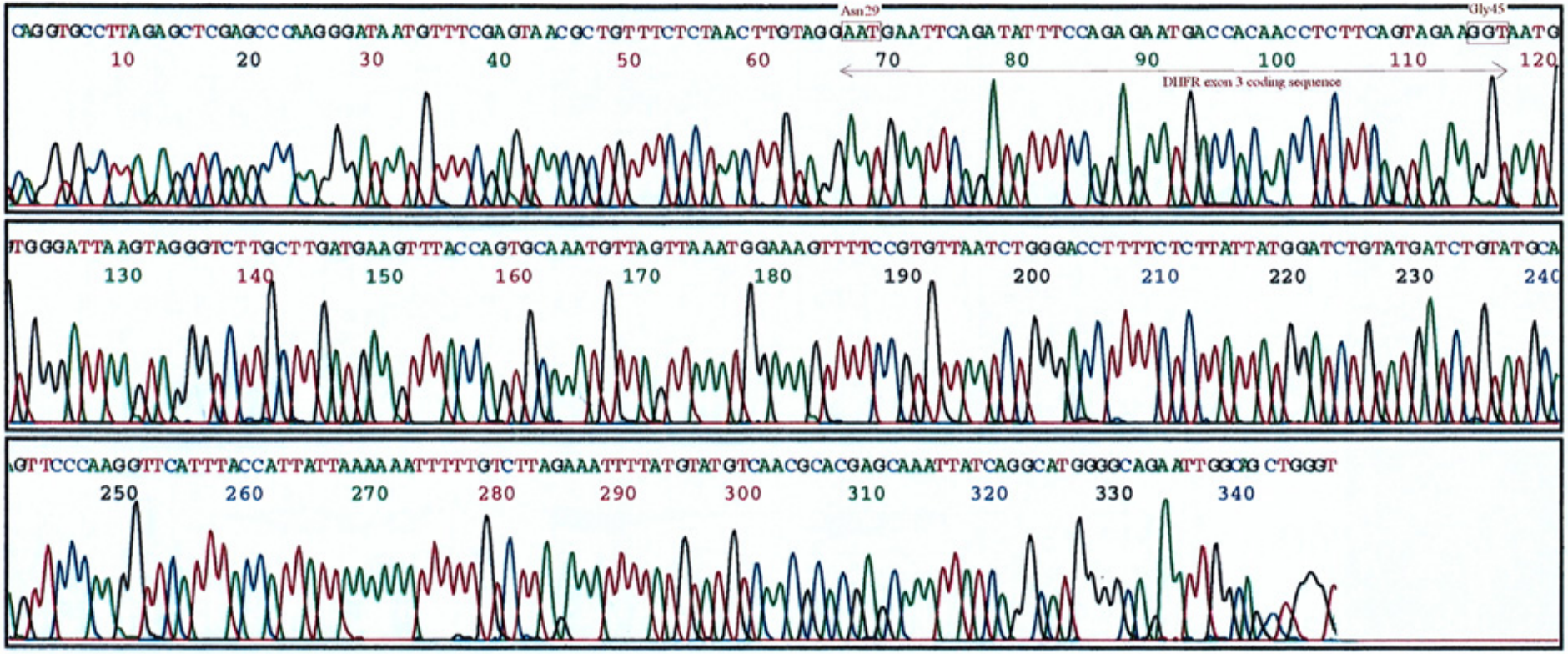
88
cccagccc tggagaaaac acctcacccc tacccacagc gctccgtttg tcaggtgcct tagagctcga gcccaaggga taatgtttcg agtaacgctg tttctctaac ttgtaggAAT

91 101 111 121 131 138
GAATTCAGAT ATTTCCAGAG AATGACCACA ACCTCTTCAG TAGAAGGTaa tgtgggatta agtagggtct tgcttgatga agtttaccag tgcaaatgtt

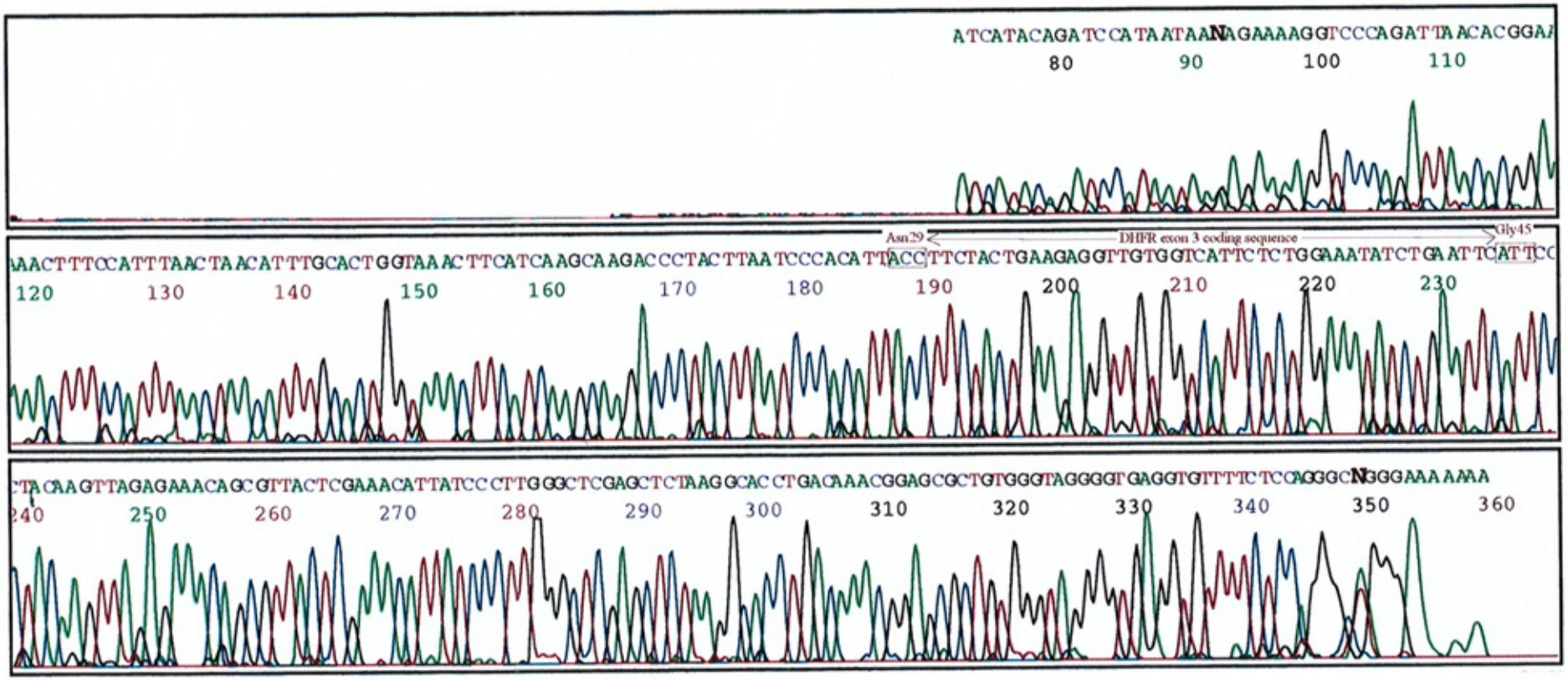
agttaaatgg aaagttttcc gtgttaatct gggacctttt ctcttattat ggatctgtat gatctgtatg cagttcccaa ggttcattta ccattattaa aaaatttttg tcttagaaat tttatgtatg tcaacgcacg

agcaaattat caggcatggg gcagaattgg cagctgggt

Appendix 8.2. Sequencing results from DHFR exon 3 forward primer for NALM6 DNA. Key features are shown. Residue number is that of the active enzyme (ie., minus start codon methionine).



Appendix 8.3. Sequencing results from DHFR exon 3 reverse primer for NALM6 DNA. Key features are shown. Residue number is that of the active enzyme (ie., minus start codon methionine).



Appendix 8.4 DHPLC results for DHFR exon 3 at 54.5 and 56.5°C. Complete profiles at all three temperatures shown in green.

DIHYDROFOLATE REDUCATASE EXON 3 **345 bp**

	DHPLC at 54.5°C			DHPLC at 56.5°C			
	Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
74(L76)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
360(L279)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
189(L405)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
L308	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
L345	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
L382	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
L609	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
980(L213)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
981(L213)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
1399(L304)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
1533	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
1775	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
1785	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
2416	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
2426	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
2650	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
2865	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
3192	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
3451	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
3605	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
4684	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
5467	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	

Continued

	DHPLC at 54.5°C			DHPLC at 56.5°C			
	Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
5836	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
5862(27/9)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
6250	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
6346	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
6563	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
6652	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
7096	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
8646	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
8678	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
8839	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
9506	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
11420	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N4	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N8	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N11	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N12	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N13	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N18	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N28	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N38	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N50	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	

Continued

	DHPLC at 54.5°C			DHPLC at 56.5°C			
	Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
N52	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N56	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N65	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N72	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N74	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N77	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N81	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N83	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N84	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	

Appendix 9.1. Part of chromosome 5 amplified by DHFR exon 4 primers (showing primers, coding sequence and key features).

Key:- Exon 4 shown in capital letters. Primers shown as shaded blocks. Other key features are shown in red.
Numbers across the top show coding base number using the first base in the start codon as position 1.

agactccaca cagacggtgg ccctggccag gaattgatta ttttctcat caacgttaga agaaaacgat gttattcaag gacatgttgt atataaatac atgaaaatat atgtatattc ccaaaatata

139 141 151 161 171 181 191
cttaggcttt ctttgtgatt ttataggtAA ACAGAATCTG GTGATTATGG GTAAGAAGAC CTGGTTCTCC ATTCCTGAGA AGAATCGACC

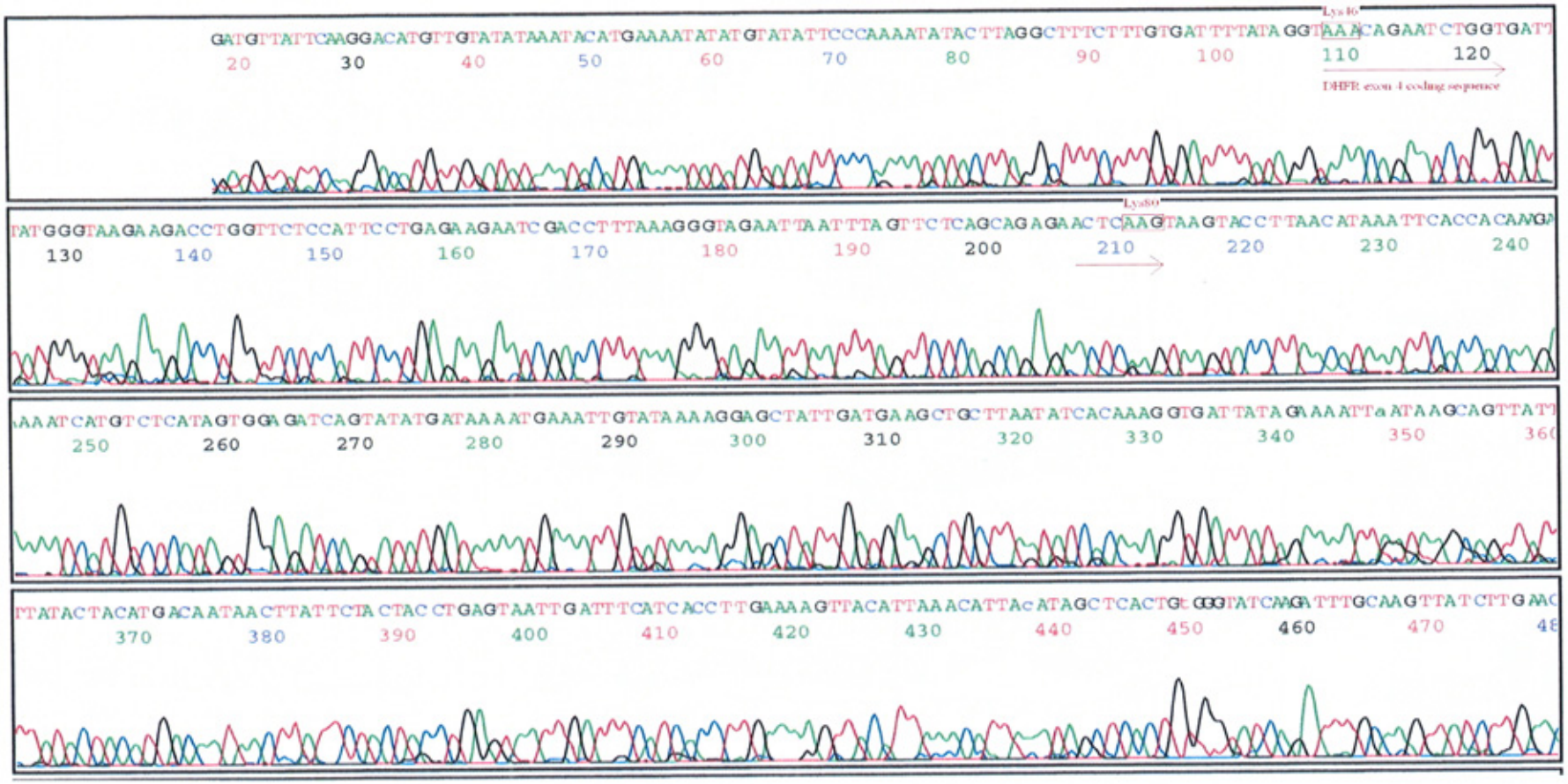
201 211 221 231 243
TTTAAAGGGT AGAATTAATT TAGTTCTCAG CAGAGAACTC AAGtaagtac cttaacataa attcaccaca agaaaatcat gtctcatagt ggagatcagt

atatgataaa atgaaattgt ataaaaggag ctattgatga agctgcttaa tatcacaaag gtgattatag aaaattaata agcagttatt tatactacat gacaataact tattctacta cctgagtaat

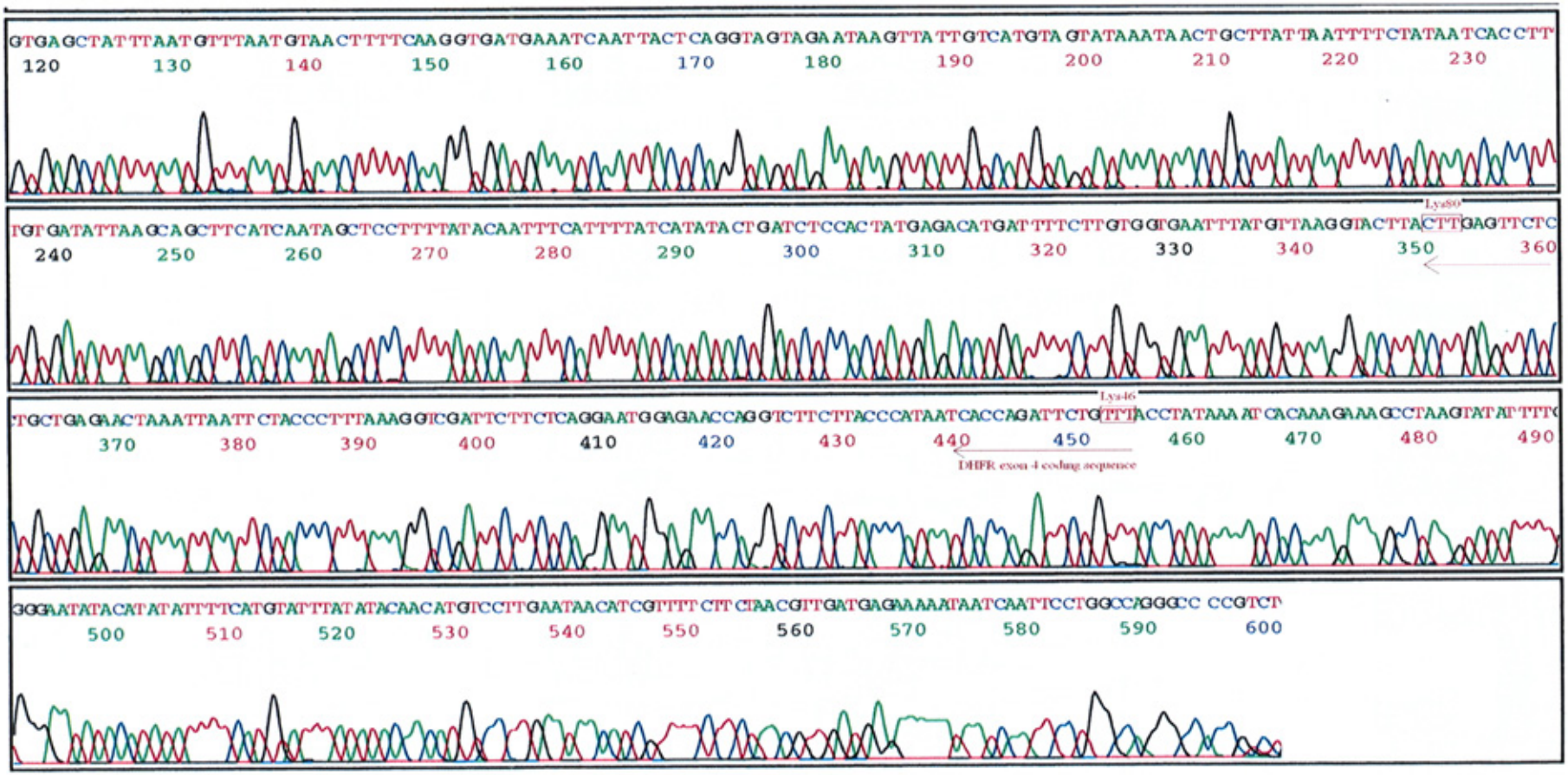
tgatttcatc accttgaaaa gttacattaa acattaaata gctcactgtg gtatcagatt tgcaagttat cttgaaccat cattcagtag gcccaaatat aaattataca gtgtaatcaa taacctgtag

tttcacataa agtcaaaaca actatttttt gggacttgct aaaattaagt gatgatcccc tttctggacc caaccct

Appendix 9.2. Sequencing results from DHFR exon 4 forward primer for NALM6 DNA. Key features are shown. Residue number is that of the active enzyme (ie., minus start codon methionine).



Appendix 9.3. Sequencing results from DHFR exon 4 reverse primer for NALM6 DNA. Key features are shown. Residue number is that of the active enzyme (ie., minus start codon methionine).



Appendix 9.4 DHPLC results for DHFR exon 4 at 54°C. Complete profiles shown in green.

DIHYDROFOLATE REDUCTASE EXON 4 **658 bp**

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
74(L76)	homoduplex	homoduplex	WT	
360(L279)	homoduplex	homoduplex	WT	
189(L405)	heteroduplex		heterozygous	
L308	homoduplex	homoduplex	WT	
L345	homoduplex	homoduplex	WT	
L382	homoduplex	homoduplex	WT	
L609	homoduplex	homoduplex	WT	
980(L213)	homoduplex	homoduplex	WT	
981(L213)	homoduplex	homoduplex	WT	
1399(L304)	homoduplex	homoduplex	WT	
1533	homoduplex	homoduplex	WT	
1775	homoduplex	homoduplex	WT	
1785	homoduplex	homoduplex	WT	
2416	homoduplex	homoduplex	WT	
2426	homoduplex	homoduplex	WT	
2650	homoduplex	homoduplex	WT	
2865	homoduplex	homoduplex	WT	
3192	homoduplex	homoduplex	WT	
3451	homoduplex	heteroduplex	Not WT	
3451	homoduplex	heteroduplex	Not WT	
3605	homoduplex	homoduplex	WT	
4684	heteroduplex		heterozygous	
5467	homoduplex	homoduplex	WT	
5836	homoduplex	homoduplex	WT	
5862(31/7)	homoduplex	homoduplex	WT	

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
6250	heteroduplex		heterozygous	
6563	homoduplex	homoduplex	WT	
6652	homoduplex	homoduplex	WT	
7096	homoduplex	homoduplex	WT	
8646	homoduplex	homoduplex	WT	
8678	homoduplex	homoduplex	WT	
9506	heteroduplex		heterozygous	
11420	homoduplex	heteroduplex	Not WT	
N4	homoduplex	homoduplex	WT	
N8	homoduplex	homoduplex	WT	
N11	homoduplex	homoduplex	WT	
N12	homoduplex	heteroduplex	Not WT	
N13	homoduplex	homoduplex	WT	
N18	homoduplex	homoduplex	WT	
N28	homoduplex	heteroduplex	Not WT	
N38	homoduplex	homoduplex	WT	
N50	homoduplex	homoduplex	WT	
N52	homoduplex	homoduplex	WT	
N56	homoduplex	empty well		
N65	homoduplex	homoduplex	WT	
N72	heteroduplex		heterozygous	
N74	homoduplex	homoduplex	WT	
N77	heteroduplex		heterozygous	
N83	homoduplex	homoduplex	WT	
N84	homoduplex	homoduplex	WT	

Appendix 10.1. Part of chromosome 5 amplified by DHFR exon 5 primers (showing primers, coding sequence and key features).

Key:- Exon 5 shown in capital letters. Primers shown as shaded blocks. Other key features are shown in red.
Numbers across the top show coding base number using the first base in the start codon as position 1.

ctgatgt taagtgcttt ttgttgaaag gattcatatt tatgatgatt ttagttctat agtcactgca tcttagtcag attatttgac tagtatggat ggattttaca aacattttgt ggaattcttt ttttaacctt

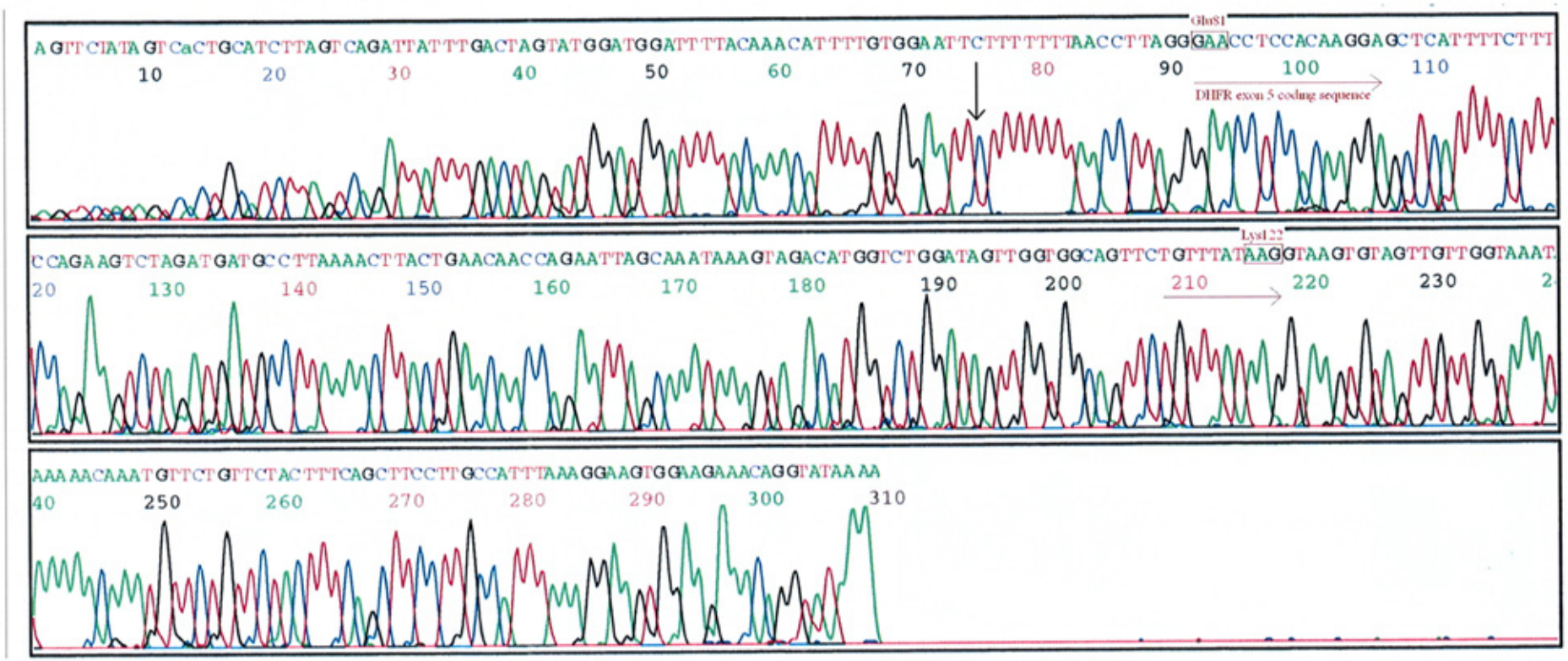
Jurkat T/G difference
Intron 4-5 base 11279

K562 T/C difference
intron 4-5 base 11362

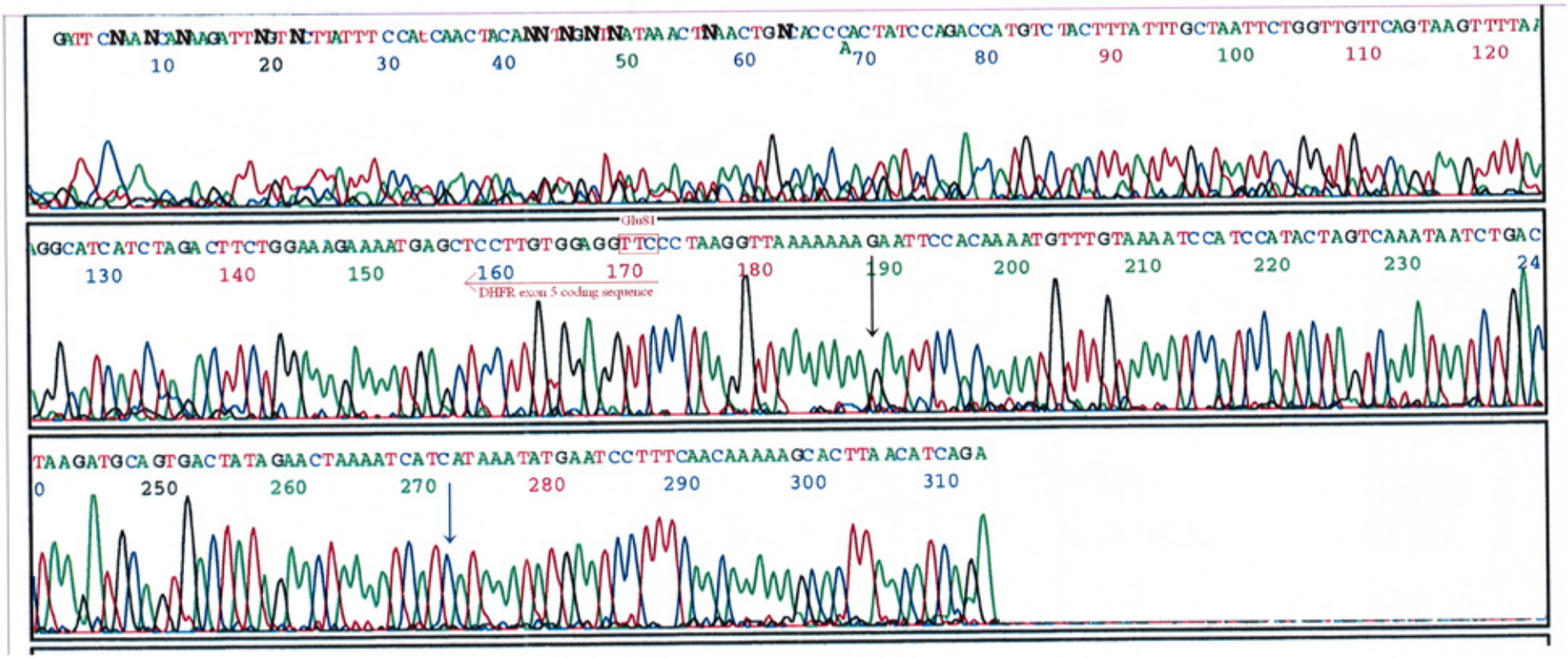
```
    244         251         261         271         281         291         301         311
aggGAACCTC CACAAGGAGC TCATTTTCTT TCCAGAAGTC TAGATGATGC CTTAAAACTT ACTGAACAAC CAGAATTAGC

321         331         341         351         361         369
AAATAAAGTA GACATGGTCT GGATAGTTGG TGGCAGTTCT GTTTATAAGg taagtgtagt tgttggtaaa taaaaacaaa tgttctgttc tactttcagc
```

ttccttgcca tttaaaggaa gtggaagaaa caggtataa

Appendix 10.2. Sequencing results from DHFR exon 5 forward primer for NALM6 DNA.
Key features are shown. Residue number is that of the active enzyme (ie., minus start codon methionine). ↓Arrow indicates position of difference between NALM6 and K562 (base 11362 intron 4-5).



Appendix 10.3. Sequencing results from DHFR exon 5 reverse primer for NALM6 DNA. Key features are shown.
Residue number is that of the active enzyme (ie., minus start codon methionine).
↓Arrow indicates position of difference between NALM6 and K562 (base 11362 intron 4-5). Blue↓Arrow indicates position of difference between NALM6 and Jurkat (base 11279 intron 4-5).



Appendix 10.4 DHPLC results for DHFR exon 5 at 54.7°C. Complete profiles shown in green. Samples with 2 peak heteroduplex patterns have been scored as heterozygote A/G at position intron 4-5 base 11362. Those with 4 peak heteroduplex patterns have no confirmed sequence change to allow genotyping.

DIHYDROFOLATE REDUCTASE EXON 5 **658 bp**

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
74(L76)	homoduplex	homoduplex	WT	
360(L279)	Heteroduplex (4)			
189(L405)	homoduplex	homoduplex	WT	
L609	Heteroduplex (2)		A/G intron	
L382	Heteroduplex (4)			
L345	Heteroduplex (4)			
L308	homoduplex	homoduplex	WT	
980(L213)	homoduplex	homoduplex	WT	
981(L213)	homoduplex	homoduplex	WT	
1399(L304)	homoduplex	homoduplex	WT	
1533	Heteroduplex (4)			8/9/2007
1775	Heteroduplex (4)			
1785	homoduplex	homoduplex	WT	
2416	Heteroduplex (2)		A/G intron	
2426	homoduplex	homoduplex	WT	
2650	homoduplex	homoduplex	WT	
2865	Heteroduplex (4)			
3192	Heteroduplex (4)			
3451	homoduplex	homoduplex	WT	
3605	homoduplex	homoduplex	WT	
4684	Heteroduplex (4)			
5467	Heteroduplex (4)			
5836	homoduplex	homoduplex	WT	
5862(31/7)	Heteroduplex (2)		A/G intron	

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced
5862(27/9)	Heteroduplex (2)		A/G intron	8/9/2007
6250	homoduplex	homoduplex	WT	
6346	homoduplex	homoduplex	WT	
6563	homoduplex	homoduplex	WT	
6652	Heteroduplex (2)		A/G intron	
7096	Heteroduplex (4)			8/9/2007
8646	homoduplex	homoduplex	WT	
8678	Heteroduplex (2)		A/G intron	8/9/2007
8839	homoduplex	homoduplex	WT	
9506	homoduplex	homoduplex	WT	
11420	homoduplex	homoduplex	WT	
N11	homoduplex	homoduplex	WT	
N12	homoduplex	homoduplex	WT	
N13	Heteroduplex (4)			
N18	Heteroduplex (2)		A/G intron	
N28	Heteroduplex (2)		A/G intron	
N31	homoduplex	homoduplex	WT	
N38	homoduplex	homoduplex	WT	
N50	Heteroduplex (4)			
N52	Heteroduplex (2)		A/G intron	
N56	homoduplex	homoduplex	WT	
N65	homoduplex	homoduplex	WT	
N72	homoduplex	homoduplex	WT	
N74	Heteroduplex (4)			

Continued Appendix 10.4

Sample #	WAVE result	WAVE result	Genotype	Date sequenced
N4	homoduplex	?Het	Rpt	
N8	homoduplex	homoduplex	WT	
N11	homoduplex	homoduplex	WT	
N12	homoduplex	homoduplex	WT	
N13	Heteroduplex (4)			
N18	Heteroduplex (2)		A/G intron	
N28	Heteroduplex (2)		A/G intron	
N31	homoduplex	homoduplex	WT	
N38	homoduplex	homoduplex	WT	
N50	Heteroduplex (4)			
N52	Heteroduplex (2)		A/G intron	
N56	homoduplex	homoduplex	WT	
N65	homoduplex	homoduplex	WT	
N72	homoduplex	homoduplex	WT	
N74	Heteroduplex (4)			
N77	homoduplex	homoduplex	WT	
N83	Heteroduplex (4)			
N84	homoduplex	homoduplex	WT	

Appendix 10.5 χ^2-test to determine if the occurrence of homoduplex or heteroduplex is significantly different between the relapse and normal groups.

	Observed			Expected		
Chromatogram	Homoduplex	Heteroduplex	Total	Homoduplex	Heteroduplex	Total
Normal	11	7	18	10.2	7.8	18
Relapse	19	16	35	19.8	15.2	35
Total	30	23	53	30	23	53

$\chi^2 = 0.63$. At 1 degree of freedom this value is $P>0.5$ and therefore null hypothesis (no difference between observed and expected) cannot be rejected.

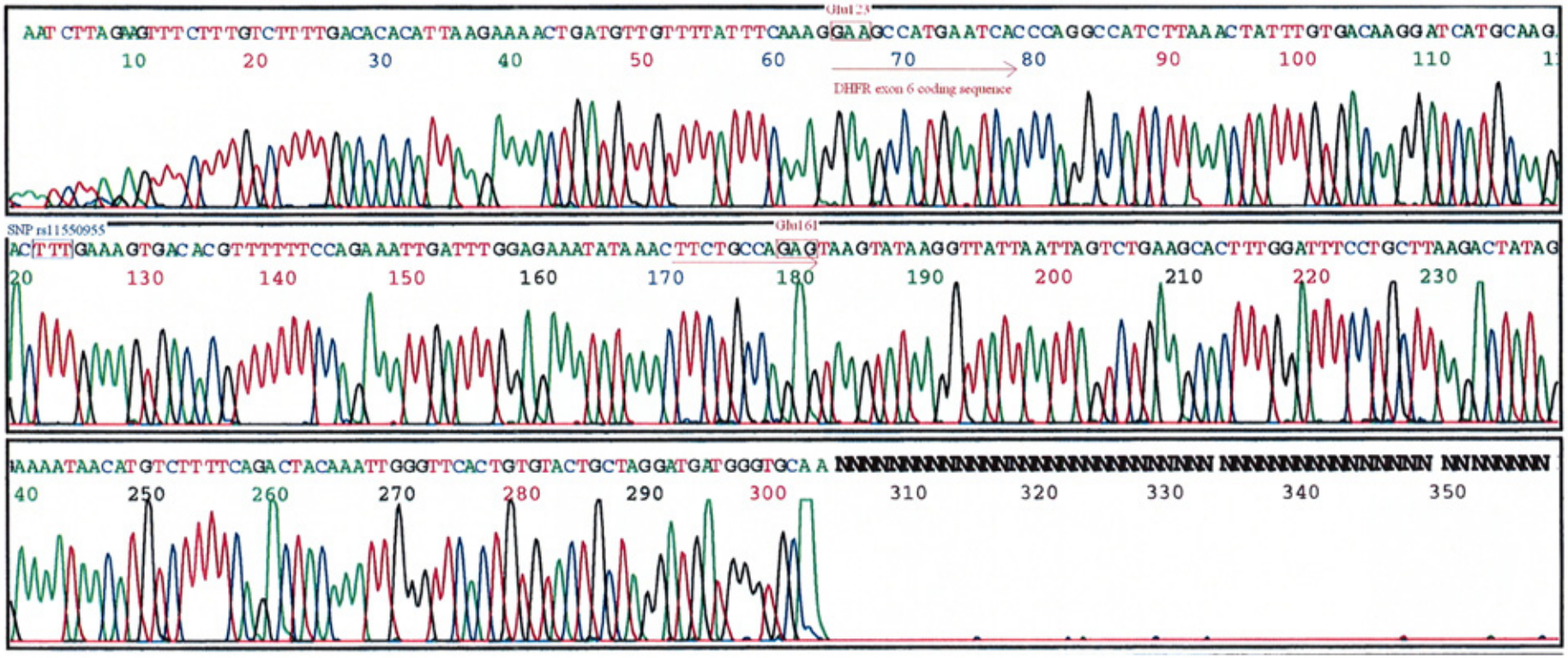
Therefore there is no significant difference between occurrence of homoduplex and heteroduplex patterns between the normal and relapse data.

Appendix 11.1. Part of chromosome 5 amplified by DHFR exon 6 primers (showing primers, coding sequence and key features).

Key:- Exon 6 shown in capital letters. Primers shown as shaded blocks. Other key features are shown in red.
Numbers across the top show coding base number using the first base in the start codon as position 1.

```
                                                                                                  370 371
ggca gcaccaagca tatttttaat acctagtata aacttaacta aacataaatc ttagagtttc tttgtctttt gacacacatt aagaaaactg atgttgtttt atttcaaagG AAGCCATGAA

381                 391                 401                 411                 421                 431                 441                 451
TCACCCAGGC CATCTTAAAC TATTTGTGAC AAGGATCATG CAAGACTTTG AAAGTGACAC GTTTTTTCCA GAAATTGATT
                                                                  T429C
                                                                  SNP rs11550955
461                 471                 481     486
TGGAGAAATA TAAACTTCTG CCAGAGtaag tataaggtta ttaattagtc tgaagcactt tggatttcct gcttaagact atagaaaata acatgtcttt tcagactaca aattgggttc

actgtgtact gctaggatga tgggtgc
```

Appendix 11.2. Sequencing results from DHFR exon 6 forward primer for NALM6 DNA. Poor results from reverse primers. Key features are shown. Residue number is that of the active enzyme (ie., minus start codon methionine).


Glu125
AAT CTTAGAAGTTTCTTTGTCTTTTGACACACATTAAGAAAACTGATGTTGTTTTATTTCAAAGGAAGCCATGAATCACCCAGGCCATCTTAAACTATTTGTGACAAGGATCATGCAAG
10 20 30 40 50 60 70 80 90 100 110
DHFR exon 6 coding sequence
SNP rs11550955
Glu161
AC[T TT]GAAAGTGACACGTTTTTTTCCAGAAATTGATTTGGAGAAATATAAACTTCTGCCAGAGTAAGTATAAGGTTATTAATTAGTCTGAAGCACTTTGGATTTCCTGCTTAAGACTATAG
20 130 140 150 160 170 180 190 200 210 220 230
AAAATAACATGTCTTTTCAGACTACAAATTGGGTTCACTGTGTACTGCTAGGATGATGGGTGCAA NNNN

Appendix 11.3 DHPLC results for DHFR exon 6 at 55.2°C. Complete profiles shown in green.

DIHYDROFOLATE REDUCTASE EXON 6 **658 bp**

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced	Sample type
74(L76)	homoduplex	homoduplex	WT		
360(L279)	homoduplex	homoduplex	WT		
392	homoduplex	homoduplex	WT		
L609	heteroduplex	homoduplex	WT		
L382	homoduplex	homoduplex	WT		
L345	homoduplex	homoduplex	WT		
L308	homoduplex	homoduplex	WT		
189(L405)	homoduplex	homoduplex	WT		
980(L213)	homoduplex	homoduplex	WT		
981(L213)	homoduplex	homoduplex	WT		
1399(L304)	homoduplex	homoduplex	WT		
1533	homoduplex	homoduplex	WT		
1566	homoduplex	homoduplex	WT		
1775	homoduplex	homoduplex	WT		
1785	homoduplex	homoduplex	WT		
2416	homoduplex	homoduplex	WT		
2426	homoduplex	homoduplex	WT		
2650	homoduplex	homoduplex	WT		
2865	homoduplex	homoduplex	WT		
3192	homoduplex	homoduplex	WT		
3451	homoduplex	homoduplex	WT		
3605	homoduplex	homoduplex	WT		
4684	homoduplex	homoduplex	WT		
5467	homoduplex	homoduplex	WT		
5836	homoduplex	homoduplex	WT		

Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced	Sample type
5862(31/7)	homoduplex	homoduplex	WT		
5862(27/9)	homoduplex	homoduplex	WT		
6250	homoduplex	homoduplex	WT		
6346	homoduplex	homoduplex	WT		
6563	homoduplex	homoduplex	WT		
6652	homoduplex	homoduplex	WT		
7096	homoduplex	homoduplex	WT		
8646	homoduplex	homoduplex	WT		
8678	homoduplex	homoduplex	WT		
8839	homoduplex	homoduplex	WT		
9506	homoduplex	homoduplex	WT		
11420	homoduplex	homoduplex	WT		

Continued

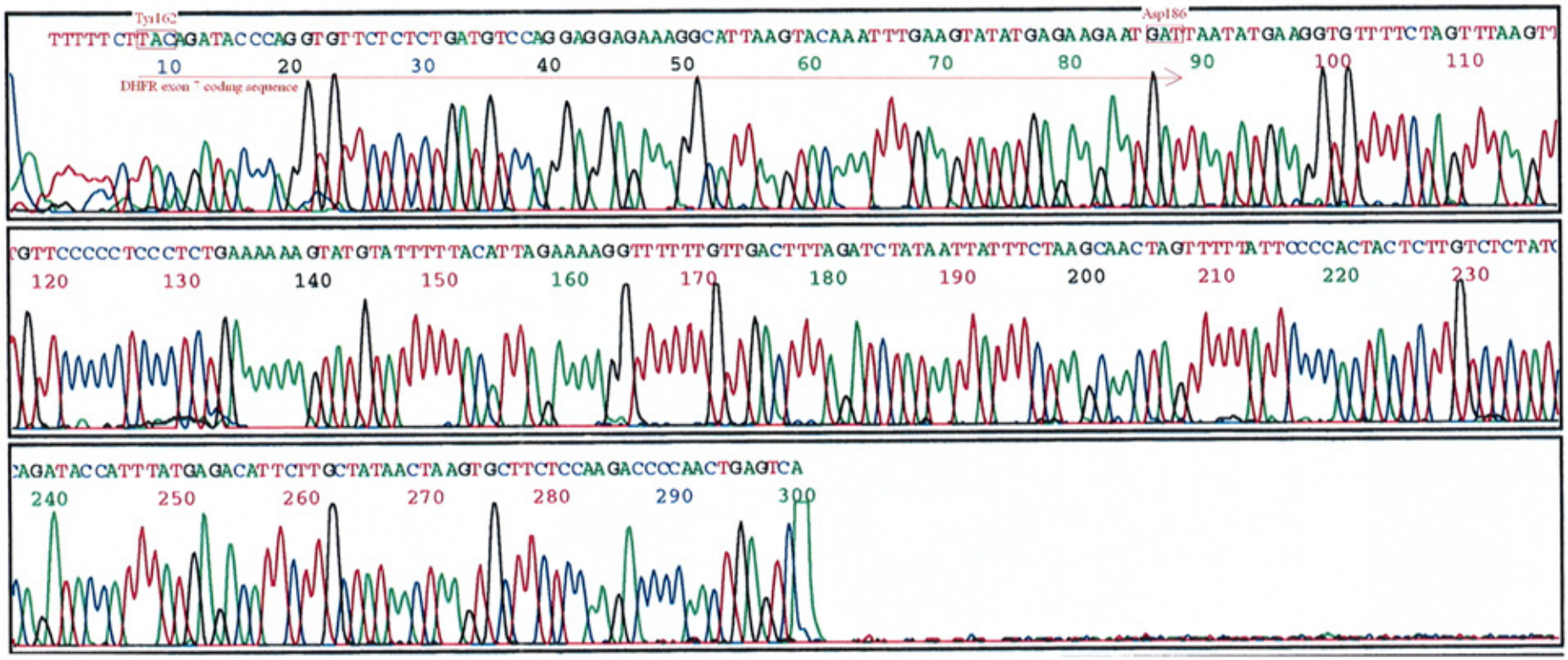
Sample #	Original WAVE result	Spiked WAVE result	Genotype	Date sequenced	Sample type
N4	homoduplex	homoduplex	WT		
N11	homoduplex	homoduplex	WT		
N12	homoduplex	homoduplex	WT		
N13	homoduplex	homoduplex	WT		
N18	homoduplex	homoduplex	WT		
N28	homoduplex	homoduplex	WT		
N31	homoduplex	homoduplex	WT		
N38	homoduplex	homoduplex	WT		
N50	homoduplex	homoduplex	WT		
N52	homoduplex	homoduplex	WT		
N56	homoduplex	homoduplex	WT		
N65	homoduplex	homoduplex	WT		
N72	homoduplex	homoduplex	WT		
N74	homoduplex	homoduplex	WT		
N77	homoduplex	homoduplex	WT		
N83	homoduplex	homoduplex	WT		
N84	homoduplex	homoduplex	WT		

Appendix 12.1. Part of chromosome 5 amplified by DHFR exon 7 primers (showing primers, coding sequence and key features).

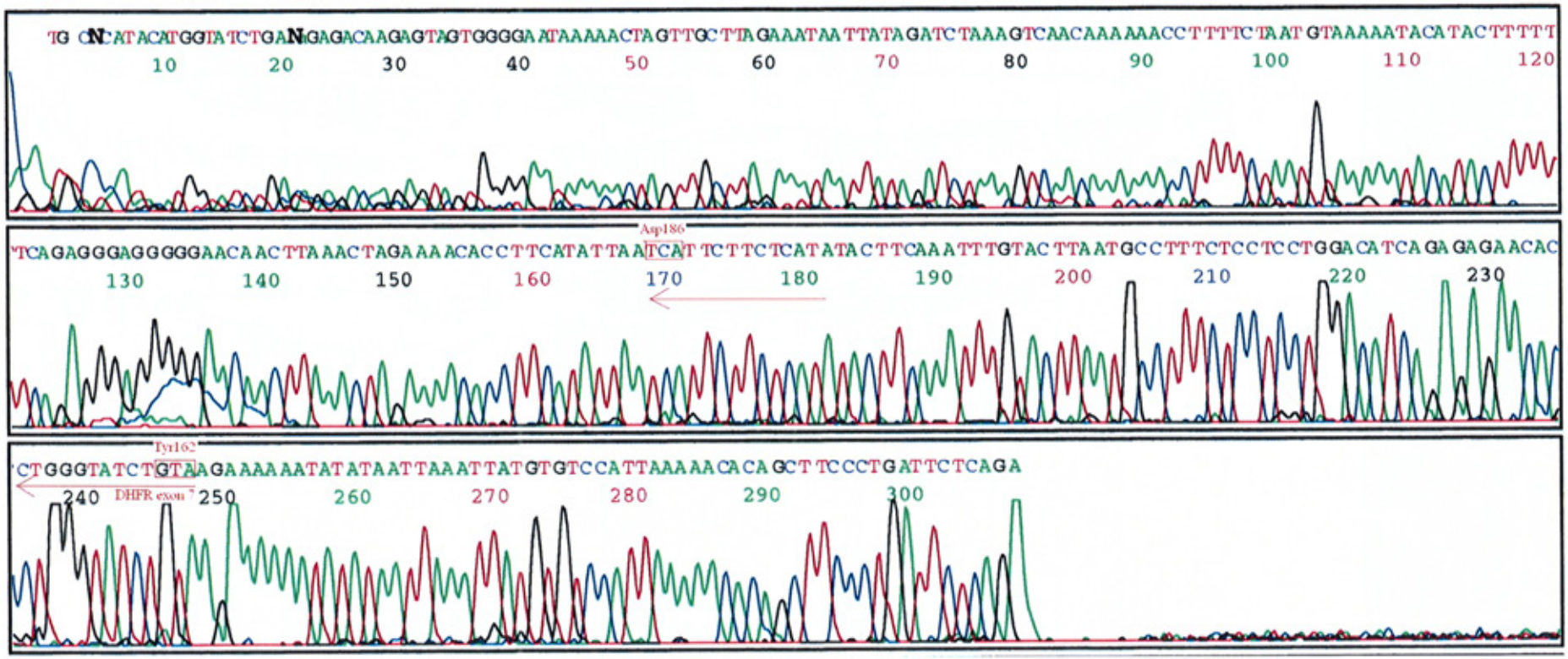
Key:- Exon 7 shown in capital letters. Primers shown as shaded blocks. Other key features are shown in red.
Numbers across the top show coding base number using the first base in the start codon as position 1.

```
                                                                                                  487       491                 501                 511                 521
ctgagaatc agggaagctg tgtttttaat ggacacataa tttaattata tatttttct tacagaTACC CAGGTGTTCT CTCTGATGTC CAGGAGGAGA AAGGCATTAA

531                 541                 551                 561
GTACAAATTT GAAGTATATG AGAAGAATGA Ttaatatgaa ggtgttttct agtttaagtt gttccccctc cctctgaaaa aagtatgtat ttttacatta gaaaaggttt tttgttgact

ttagatctat aattatttct aagcaactag tttttattcc ccactactct tgtctctatc agataccatt tatgagacat tcttgctata actaagtgct tctccaagac cccaactgag tc
```

Appendix 12.2. Sequencing results from DHFR exon 7 forward for NALM6 DNA.
Key features are shown. Residue number is that of the active enzyme (ie., minus start codon methionine).



Appendix 12.3. Sequencing results from DHFR exon 7 reverse primer for NALM6 DNA.
Key features are shown. Residue number is that of the active enzyme (ie., minus start codon methionine).



Appendix 12.4 DHPLC results for DHFR exon 7 at 53.6 and 54.6°C. Complete profiles at all three temperatures shown in green.

DIHYDROFOLATE REDUCATASE EXON 7 **345 bp**

	DHPLC at 53.6°C			DHPLC at 54.6°C			
	Original	**Spiked**		**Original**	**Spiked**		
Sample #	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**Date sequenced**
74(L76)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
360(L279)	heteroduplex		heterozygous	heteroduplex		heterozygous	
L609	heteroduplex			heteroduplex			
L382	heteroduplex			heteroduplex			
L343	heteroduplex			heteroduplex			
L308	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
189(L405)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
980(L213)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
981(L213)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
1399(L304)	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
1533	heteroduplex		heterozygous	heteroduplex		heterozygous	
1775	heteroduplex		heterozygous	heteroduplex		heterozygous	
1785	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
2416	heteroduplex			heteroduplex			
2426	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
2650	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
2865	heteroduplex		heterozygous	heteroduplex		heterozygous	
3192	heteroduplex			heteroduplex			
3451	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
3605	heteroduplex	homoduplex	WT	heteroduplex	homoduplex	WT	
4684	heteroduplex		heterozygous	heteroduplex		heterozygous	
5467	heteroduplex		heterozygous	heteroduplex		heterozygous	

Continued

	DHPLC at 54°C			DHPLC at 66°C			
	Original	Spiked		Original	Spiked		
Sample #	WAVE result	WAVE result	Genotype	WAVE result	WAVE result	Genotype	Date sequenced
5836	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
5862(31/7)	heteroduplex		heterozygous	heteroduplex		heterozygous	
5862(27/9)	heteroduplex		heterozygous	heteroduplex		heterozygous	
6250	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
6346	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
6563	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
6652	heteroduplex			heteroduplex			
7096	heteroduplex		heterozygous	heteroduplex		heterozygous	
8646	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
8678	heteroduplex		heterozygous	heteroduplex		heterozygous	
8839	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
9506	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
11420	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N4	homoduplex	?	?	homoduplex	?	?	
N8	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N11	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N13	homoduplex	?	?	?hetero	?	?	
N18	heteroduplex		heterozygous	heteroduplex		heterozygous	
N28	heteroduplex		heterozygous	heteroduplex		heterozygous	
N38	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N50	heteroduplex		heterozygous	heteroduplex		heterozygous	
N52	heteroduplex		heterozygous	heteroduplex		heterozygous	

Continued

	DHPLC at 54°C			DHPLC at 55°C			
	Original	**Spiked**		**Original**	**Spiked**		
Sample #	**WAVE result**	**WAVE result**	**Genotype**	**WAVE result**	**WAVE result**	**Genotype**	**Date sequenced**
N65	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N72	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N74	heteroduplex		heterozygous	heteroduplex		heterozygous	
N77	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N81	homoduplex	homoduplex	WT	homoduplex	homoduplex	WT	
N83	heteroduplex		heterozygous	heteroduplex		heterozygous	

Denaturing high-performance liquid chromatography: the tool of choice for detecting single nucleotide polymorphisms

Of the techniques currently available to detect single nucleotide polymorphisms, denaturing high-performance liquid chromatography (DHPLC) has proved to be both the fastest and simplest. Here, Geoffrey Bosson looks at the awesome task of screening the genome

Now that the complete human genome is known, it is clear that the 30,000 genes do not account for the phenotypic diversity of the human population. The biodiversity must be due to minor alterations in the genome (single nucleotide polymorphisms [SNPs]), that either go unnoticed or produce an evolutionary advantage or disadvantage. It is the latter that genetics as a science is focused on, because it is these that result in disease.

The technology for determining the human genome advanced greatly in the 1990s but is still not suitable for screening large populations to find SNPs. Of the techniques developed, denaturing high-performance liquid chromatography (DHPLC) is proving both the fastest and simplest, and has an acceptable diagnostic accuracy. Genes found to contain SNPs can then be sequenced to determine the exact base sequence change and, by comparing it with the wild-type gene, or in transfection studies, elucidate what effect this will have when transcribed and translated.

DNA and genes

With a few exceptions, DNA is responsible for the coding of life. The DNA molecule is made up of two complementary strands, each composed of a sequence of nucleic acids arranged along a backbone of deoxyribose and phosphate.[1] Sequences of nucleic acid bases along a single strand are grouped into regions called genes. When a gene is converted into a complementary piece of mRNA (ie transcribed) and then converted into a sequence of amino acids (ie translated) a specific protein is produced.

The two complementary strands of DNA run in opposite directions (Figure 1). This means that one is aligned with its available deoxyribose (ie that containing the free hydroxyl on carbon-5) at one end and the free phosphate (attached to carbon-3 of the ribose) at the other end (Figure 2), and is referred to as the 5'→3' strand. The other strand runs in the 3'→5' direction. The linkage between the two strands is via the nucleic acid bases that form bridges (sometimes referred to as rungs because of the ladder-like appearance of DNA). The bases normal-

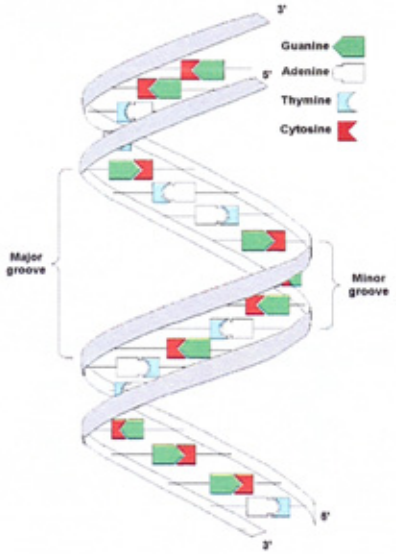


Fig 1. Diagrammatic representation of double stranded DNA showing the ladder-like appearance of the helical structure

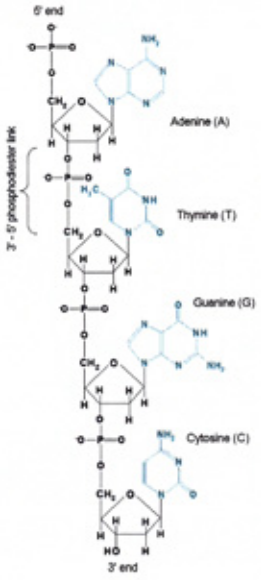


Fig 2. Deoxyribose phosphate backbone (bold) and the four possible nucleotides (grey)

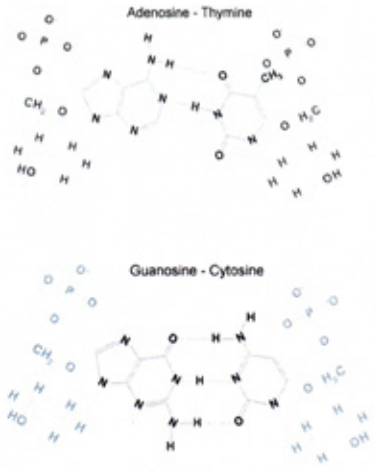


Fig 3. Base pairing of adenosine with thymine and guanosine with cytosine

ly only pair with one specific counterpart, so that adenine is paired to thymine, guanine to cytosine etc (Figure 3). The paired bases are linked via hydrogen bonds, two between A-T and three between C-G.[2] This gives the latter a greater stability.

The other significant point about the base pairing is that normally a single ring pyridine (cytosine or thymine) is paired with a two-ring purine (adenine or guanine). This ensures that the distance between the two strands is normally kept constant.

The whole human genome is spread across the 22 paired autosomal chromosomes plus the X and Y chromosomes. Regions of a chromosome that code for a specific protein are referred to as genes. In addition to these protein-coding genes, there are some that act as signals to switch on (operons) or switch off (suppressors) the transcription of a specific gene. Much of the work on genes has been done during investigation of cancer and the three-letter code used to identify a gene usually reflects this. Hence, genes that when switched on cause cancerous change are often called oncogenes (eg *ras* [rat sarcoma]; *erb* [erythroblastosis virus]).[3] Repressor genes normally exist in a switched-on state and are responsible for suppressing transcription of certain genes, but a change in a single base can result in failure of this suppression and lead to cancer. A major suppressor gene is p53, which is involved in protection of existing DNA and its repair when damaged.[4]

DNA damage

Alterations in the base sequence of DNA can occur while in the duplex form (eg formation of a cyclobutyl ring between adjacent thymine residues, as shown in Figure 4), but the majority occur during replication, when complementary daughter strands are formed using the original strand as a template. If these errors are not repaired they are further replicated and make the mutation significant to the whole organism.

Mutations of the DNA can be classed as either point mutations, where one base replaces another (eg transitions that are replaced by a similar nucleotide [C→T or T→C, A→G or G→A]; transversions that are replaced by different nucleotide [T→A or T→G, C→A or C→G, A→T or A→C, G→T or G→C]) or insertions/deletions where there is the loss or addition of a base or bases.

The causes of such mutations include the presence of nucleotide analogues that are mistakenly incorporated into the DNA and then paired with a different base when replicated (eg 5-bromouracil resembles thymine, but prefers to pair with adenine instead of guanine, therefore causing an A-T pair transition to a G-C, as shown in Figure 5).

Chemical modifications, such as deamination, alkylation and methylation of nucleic acids will cause point mutations. Dyes like acridine orange can fit between consecutive base pairs causing stretching of the strand. When this region replicates, the dye effectively doubles the intercalation distance and allows the insertion of a free base.[3]

Ionising radiation is capable of transferring energy to the DNA causing disruption of bonds, while energy from UV light is capable of forming covalent linkage between adjacent pyrimidines (T-T, C-C, C-T) forming dimers (Figure 4) that are misread during replication.

DNA repair

The majority of mutations occur in regions of the genome that are redundant and therefore go unnoticed. It is only when the mutation is in an active coding gene, repressor gene or oncogene that some effect is noted. The number of muta-

Table 1. Base stacking energies of all 10 possible dimer permutations[5]

C-G	-61.0
G-C	
C-G	-44.0
A-T	
C-G	-41.0
T-A	
G-C	-40.5
C-G	
G-C	-34.6
G-C	
G-C	-28.4
A-T	
T-A	-27.5
A-T	
G-C	-27.5
T-A	
A-T	-22.5
A-T	
A-T	-16.0
T-A	

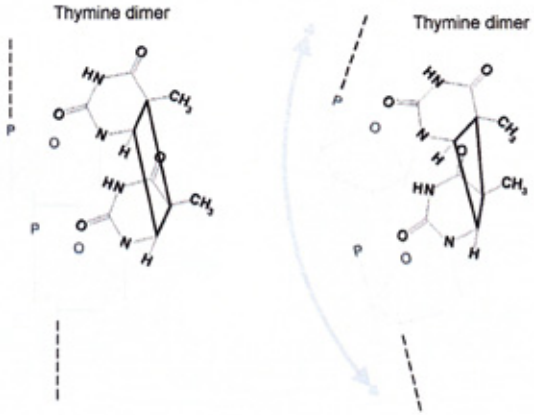


Fig 4. Consecutive thymine residues held together by a cyclobutyl ring to form a dimmer. The energy in the ring pulls the two bases together and puts an unnatural bend in the strand of DNA causing a misread in the next replication

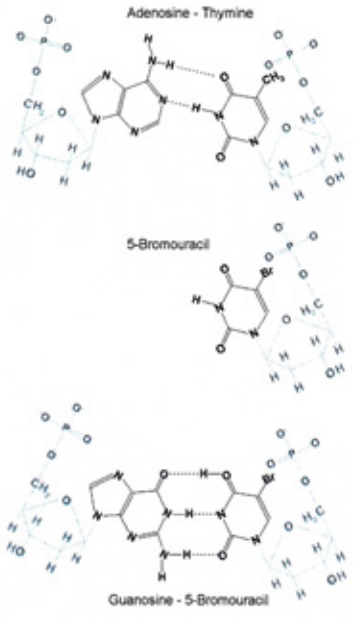


Fig 5. The original A-T pairing (top) and how 5-bromouracil (middle) can be mistaken for thymine and integrated into the daughter strand during replication. When the strand containing the 5-bromouracil is replicated next time, a molecule of guanine will pair with the 5-bromouracil in place of the original adenosine (bottom)

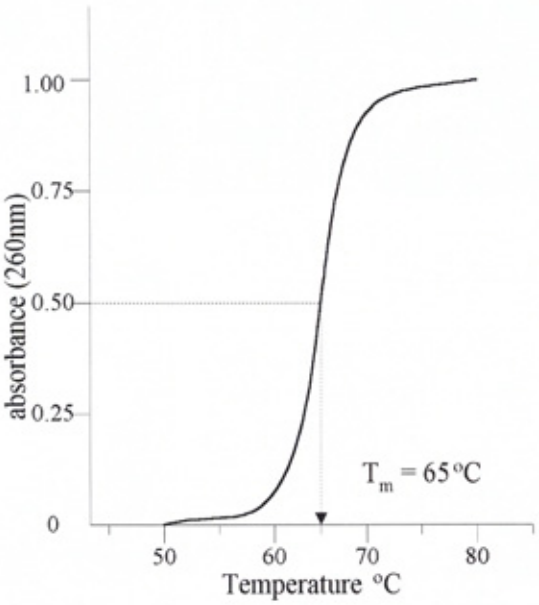


Fig 6. Typical melting curve of DNA

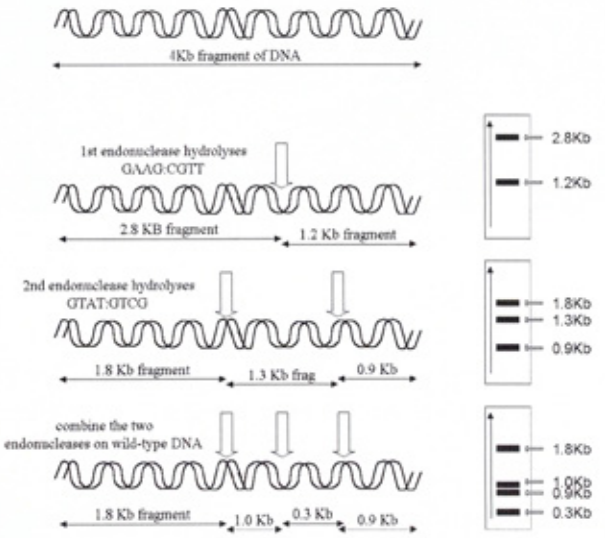


Fig 7. Diagram showing use of two restriction endonucleases to demonstrate the presence of a single nucleotide polymorphism (adapted from Voet, Voet and Pratt[10])

tions calculated to occur in a cell's life is not matched by the relatively low number of actual mutations found. This could be due to apoptosis of the mutated cell, or it could be the result of DNA repair.

The type of repair system used depends on the type of damage detected and includes base excision (removing the nucleic acid without breaking the deoxyribose phosphate backbone, alkyl transferase (removing base adducts), nucleotide excision (removing the whole nucleotide involves breaking the backbone – p53 is important at this point as it suppresses replication until repair is complete), mismatch repair (this uses a complex of proteins capable of post-replication recognition, excision and repair) and DNA strand breaks (polymerisation or the addition of adducts distort the shape of the DNA causing strand breaks that must be repaired using the undamaged paired chromosome as template).

When the damage is too great to repair, the only solution is to prevent the cell from replicating, by causing cell death (apoptosis). This is the responsibility of the BAX gene.

Significance of SNPs

The majority of work by molecular geneticists has focused on the carcinogenic effects of SNPs, but mutations are also the cause of metabolic variations. The alteration of a specific enzyme can produce a population with increased metabolism, decreased activity or even total absence of a particular reaction. The resulting inborn errors of metabolism are well documented, but there are some enzymes or isoenzymes that are responsible for metabolism detoxification. The mutations only become apparent when the cell is exposed to a particular xenobiotic. This can explain why some individuals tolerate exposure to a toxin (eg cigarette smoke) while others develop cancer.

Denaturation of DNA

Double-stranded DNA must separate into single strands for replication and transcription. This destabilisation of DNA duplexes (also known as denaturation or melting) is required for the majority of studies and techniques, and can be achieved by heating the extracted DNA. Thermal energy is sufficient to break the hydrogen bonding between paired bases.

Interestingly, the absorbance of the single strands is greater than the duplex when measured at 260nm. This means that denaturation of DNA can be followed and the absorbance against temperature plotted. Such a graph is referred to as the melting curve (Figure 6). The resulting curve is a characteristic sigmoid curve which suggests that denaturation of one base makes the surrounding bases less stable and the rate of denaturation accelerates. The temperature at which half the change in absorbance is achieved is called the melting temperature (T_m).

G-C pairing, with its three-hydrogen bonds is more stable than A-T pairs and thus a high molar G-C% produces a higher T_m. Other factors that influence the T_m are pH, ionic concentration and any base pair mismatches.

The interbase hydrogen bonding is too weak by itself to explain the stability of the DNA duplex. The association of bases forming the ring stacks have multiple weak van der Waals' interactions and those with the greatest hydrophobic groups will attract and protect each other. There are 10 possible dimer stacking permutations (Table 1) with C-G/G-C requiring the most energy to dissociate and A-T/T-A requiring the least. (See Table 1)

This variable stability is utilised by replication enzymes as the weak point at which the duplex can be denatured. Knowledge about the influence that near neighbours have on denaturation is exploited by molecular genetics in analytical methods,[6,7] including DHPLC.[8] Base pair mismatching, insertions and deletions alter the base stacking energies so that the T_m of the mutated DNA is different from the wild type.

Methods for detecting SNPs

The ultimate method for detecting and identifying mutations is to perform full sequencing, but even with the advanced automated systems currently available this is still unrealistic for large populations.[9]

Restriction mapping

Restriction endonucleases were originally obtained from bacteria, where they function to destroy invading viral DNA. The enzymes recognise specific base sequences and hydrolyse the bond at this particular point. When a specific endonuclease is added it will cut a strand of DNA into a number of fragments depending on how

often and where the base sequence is found. These can then be separated by electrophoresis and visualised as bands of different sizes. Each sample will give the same pattern of bands when the base sequence is the same. A piece of DNA with an SNP in the region normally recognised by the endonuclease will no longer be recognised and the pattern of fragments will alter. (Figure 7). Similar fragmentation can be achieved by chemical cleavage of mismatched bases (eg osmium tetroxide/piperidine attacks mismatched T bases, hydroxylamine/piperidine attacks mismatched C bases etc).

DNA probes

Use of DNA probes has started to challenge for the position of best screening method, especially with the development of DNA chip arrays.[9] These consist of a series of known oligonucleotide sequences (probes) attached to a stationary support (chip). If the DNA in an unknown sample possesses a complementary sequence to the probe, it will bind and compete with added fluorescent-labelled wild-type DNA. The result is that less fluorescent label will bind to the particular area of the chip holding the matching probe. However, if the sample DNA contains an SNP in the relevant sequence, then it is no longer complementary to the probe and cannot compete or block the labelled DNA. Thus, a highly fluorescent spot on the array indicates a mutation in this gene. The ability to attach many thousands of probes onto a single chip facilitates the simultaneous screening of numerous genes

Heteroduplex formation

If denatured DNA is allowed to cool slowly then the reannealing process ensures that matched strands (homologous) will reform perfectly into homoduplexes. However, if the sample contains an SNP/mutation in one of the alleles, then heteroduplexes (hybrids) will be formed (Figure 8). Occasionally, the mutation could affect both alleles and thus result in homoduplex formation. To check for this, the PCR amplification product is spiked with an equal amount of wild-type DNA. If the two are the same then homoduplexes will form, if different then heteroduplexes will form. Different heteroduplexes will have slightly different melting temperatures and can be differentiated by performing electrophoresis[7] or chromatographic separation[11] at a partial melting temperature. The temperature is critical and must be maintained within tight limits if reproducibility is to be obtained.

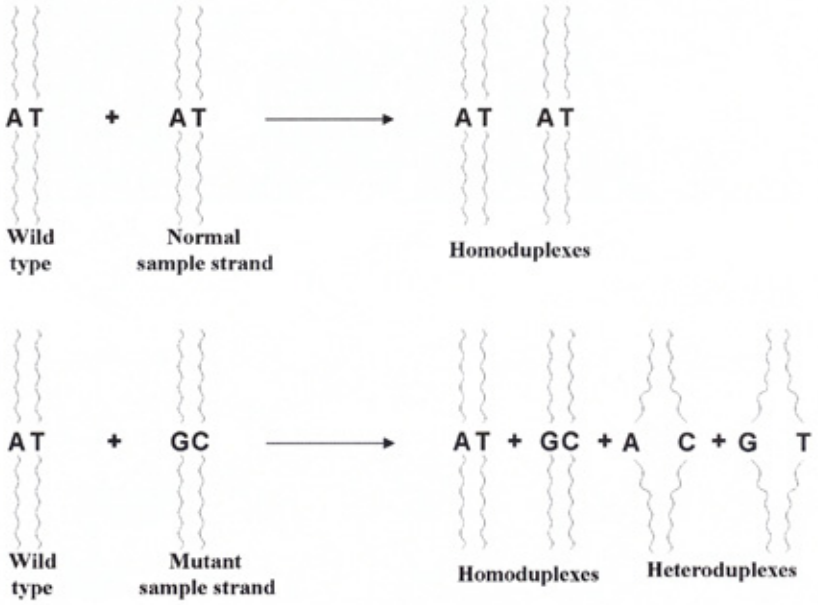


Fig 8. Diagram showing homoduplex formation when the sample strand of interest is the same as the wild type (top), and formation of four species (two homoduplexes with matching bases and two unmatched heteroduplexes) if the sample strand contains a SNP/mutation (bottom)

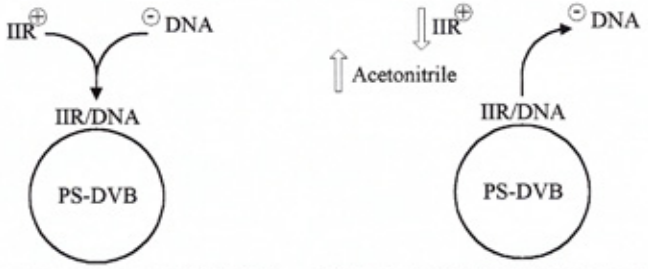


Fig 9. Ion-pairing and association with non-polar stationary phase (left) and effect of acetonitrile gradient (right)

Denaturing high performance liquid chromatography

High-performance liquid chromatography (HPLC) as been used since the 1980s to separate single-stranded DNA of different sizes,[12,13] but its use in the separation of duplex DNA was only recently achieved.[11] The performance of chromatographic separation at a specific temperature even allows separation of duplex strands of closely related molecular weight.

The HPLC system used for the separation of DNA is a reverse phase system (ie polar mobile phase and non-polar stationary phase). A method for synthesis of the non-porous polystyrene (divinylbenzene) beads for the stationary phase was described by Huber *et al.*[4] in 1993. The unique modification that allowed separation of strands of very similar size was the introduction of polyvinyl alcohol during polymeristion to give an alkylated bead.

The mobile phase contains an ion-interactive reagent (IIR) that pairs with the solute, making it non-polar.[15] The neutralised solute particle, in this case the DNA, associates with the non-polar stationary phase (shown on the left in Figure 9). Introducing a gradient to break the ion-pairing by reducing the ion-pairing buffer will cause the DNA to partition into the mobile phase. This is further enhanced by introducing acetonitrile into the mobile phase to modify its polarity. The gradient rate of change and start and end points can be selected for a specific DNA fragment.[16]

Ion selection

The positively charged ion (cation) chosen for the ion-pairing buffer needs to be able to fit into the large grove in which

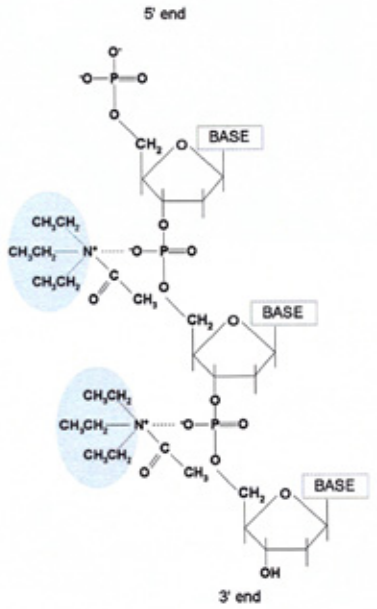


Fig 10. Interaction of positively charged triethylammonium ions with the negatively charged phosphate backbone

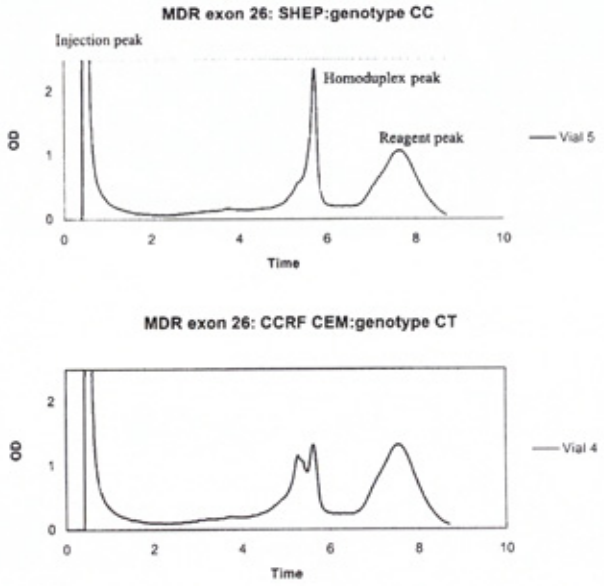


Fig 11. WAVE chromatograms showing exon 26 products from the multidrug resistance gene. (Top) CC genotype showing only a homoduplex peak; (Bottom) CT heterozygote pattern showing homoduplex and heteroduplex peaks (Reproduced with kind permission of Professor A Hall and Dr J A E Irving, Paediatric Oncology Research Unit, Newcastle University)

it binds to the negatively charged (anionic) phosphate. Tetrapropyl- and tetrabutyl-ammonium buffers are too big to fit and are of little value,[14] while tetramethyl-, triethyl- and tetraethylammonium ions in TMA, TEA and TEEA buffers, respectively, can. Any one of these cation alkyl ammonium ions is suitable.

Acetonitrile

As the fragment sizes increase, more acetonitrile is required to elute them, hence retention time is also related to fraction size.[6,14]

Temperature selection

To find the optimum temperature, an empirical approach can be applied. A heteroduplex is repeatedly injected at temperatures increasing by 3°C, starting at 50°C. At this low temperature both hetero- and homoduplexes will bind equally. Eventually a temperature will be reached at which the heteroduplexes start to melt and binding to the stationary phase is reduced, giving it a shorter retention time. Two or three peaks will be seen, depending on the polymorphism. If the temperature continues to increase then the homoduplexes start to denature and move to a shorter retention time as well.[17]

This process can be circumvented to some degree if the base sequence of the melting region is known. An algorithm is available from the University of Stanford which uses the stacking energies to predict the T_m.[18] If there is more than one suspected melting domain then the samples of PCR products need to be run at this second temperature to ensure complete SNP detection.

The WAVE

Until recently the only commercial manufacturer of a stationary phase was Trangenomic (San Jose, CA, USA), which sold columns marketed under the trade name DNASep. In 2001, Varian (Walnut Creek, CA, USA) introduced the Eclipse dsDNA column based on the same principle of separation, as part of its ProStar Helix System. Transgenomic market the DNASep column as part of an integrated HPLC system called the WAVE or, more correctly, the WAVE Nucleic Acid Fragment Analyser System Technology. It is this column and instrument that the majority of data on DHPLC and SNP screening has been produced. The DNASep column is guaranteed for 4000 injections on the WAVE system.

The instrument consists of a temperature-controlled column and includes buffers and injection valve. Thus, all critical components in the system are maintained within 0.1°C, across a range of 35-80°C. An automatic sample holder is also temperature controlled and gives fully automated injections of between 1 and 100 mL of sample from a 96-well microtitre plate.

A UV detector can be replaced by a fluorimeter to give the WAVE HS system, making it suitable for mutation screening, polymorphic marker mapping, linkage analysis etc. The ion-pair mobile phase is a triethylammonium acetate (TEAA) buffer (pH 7.0) with added 0.1 mol/L Na_2 EDTA. The elution gradient is developed using 25% acetonitrile in TEAA as buffer B. The gradient, system temperature and interpretation of results are controlled by WAVEMAKER dedicated software. This software includes a version of the temperature prediction algorithm available on the internet.[18]

An important analytical tool

Denaturing HPLC is proving to be an important analytical tool in screening the genomes of various species for SNPs and mutations. This awesome task utilises the speed of the technique, its ease of use and relative low cost. Its impact in this rapidly growing field is demonstrated by the increasing number of relevant articles found on *Medline* (two in 1997; six in 1998; eight in 1999; 30 in 2000 and 56 in 2001). As more research groups find the money to purchase the necessary instrumentation, an increasing number of papers will continue to reflect the importance of the technique. ■

REFERENCES

1 Zimmerman SB. The three-dimensional structure of DNA. *Annual Rev Biochem* 1982; **51**: 395-427.

2 Horton HR, Moran LA, Ochs RS, Rawn JD, Scrimgeour KG. *Principles of Biochemistry* Englewood Cliffs, NJ: Prentice-Hall, 1993.

3 King RJB. *Cancer Biology*. 2nd Ed. Harlow: Pearson Education, 2000

4 Gross E, Kiechle M and Arnold N. Mutation analysis of *p53* in ovarian tumors by DHPLC. *J Biochem Biophys Methods* 2001: **47**; 73-81.

5 Ornstein RL, Rein R, Breen DL and MacElroy RD. *Biopolymers* 1978; **17**: 2356.

6 Melchior (Jr) WB, Von Hippel P. Alteration of the relative stability of dA•dT and dG•dC base pairs in DNA. *Proc Nat Acad Sci USA* 1973; **70**: 298-302.

7 Ke S-H and Wartell RM. Influence of nearest neighbor sequence on the stability of base pair mismatches in long DNA: determination by temperature-gradient gel electrophoresis. *Nucleic Acids Res* 1993; **21**: 5137-43.

8 O'Donavan MC, Oefner PJ, Roberts SC *et al.* Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics* 1998; **52**: 44-9.

9 Wang DG, Lipshutz R. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998; **280**: 1077-82.

10 Voet D, Voet J, Pratt CW. *Fundamentals of Biochemistry* New York: John Wiley & Sons, 1999.

11 Oefner PJ, Underhill PA. Comparative DNA sequencing by denaturing high-performance liquid chromatography (DHPLC). *Am J Hum Genet* 1995; **57** (suppl): A266.

12 Doris PA, Oefner PJ, Chilton BS, Hayward-Lester A. Quantitative analysis if gene expression by ion-pair high-performance liquid chromatography. *J Chromatogr A* 1998; 806: 47-60.

13 Huber CG, Oefner PJ, Bonn GK. High-resolution liquid chromatography of oligonucleotides on nonporous alkylated styrene-divinylbenzene copolymers. *Anal Biochem* 1993; **212**: 351-8.

14 Huber CG, Oefner PJ, Bonn GK. Rapid and accurate sizing of DNA fragments by ion-pair chromatography on alkylated nonporous poly(styrene-divinylbenzene) particles. *Anal Chem* 1995; **67**: 578-85.

15 Robards K, Haddad PR, Jackson PE. Principles and practices of modern chromatographic methods. London: Academic Press: London, 1994: 344-51.

16 Oefner PJ. Allelic discrimination by denaturing high-performance liquid chromatography. *J Chromatogr B Biomed Appl* 2000; **739**: 345-55.

17 Giordano M, Oefner PJ, Underhill PA *et al.* Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. *Genomics* 1999; **56**: 247-53.

18 University of Stanford <http://insertion.stanford.edu/melt.html>

Geoffrey Bosson is senior lecturer in the School of Applied and Molecular Sciences,
Northumbria University, Newcastle-upon-Tyne.

Reduced folate carrier: biochemistry and molecular biology of the normal and methotrexate-resistant cell

GEOFFREY BOSSON
School of Applied and Molecular Sciences, Ellison Building, Northumbria University, Newcastle-upon-Tyne NE1 8ST, UK

Accepted: 21 March 2003

ABSTRACT

The cytotoxic drug methotrexate uses the reduced folate carrier for transport into the cell, where it inhibits key enzymes in nucleotide biosynthesis. Resistance to methotrexate can be achieved by altering the genetic code of the reduced folate carrier gene and thus change the structure and function of the protein. Our understanding of RFC structure and function is based on the information gained from studying the uptake of folates and antifolates in living cells and the application of molecular techniques to determine gene expression and genetic mutations. The aim of this essay is to explain the structure and function of the reduced folate carrier, review the molecular biology of the reduced folate carrier gene and the mutations and polymorphisms that can result in methotrexate resistance.

KEY WORDS: Drug resistance. Folic acid. Methotrexate. Molecular biology.

Introduction

The cellular synthesis of organic biomolecules involves either the joining of two or more complex molecules or simply the addition of atoms to existing precursor molecules. In such anabolic pathways, the sequential addition of single carbon atoms provides greater flexibility, and an almost infinite series of organic compounds can be produced.

Carbon atoms can be presented as either carbon dioxide (CO_2), methyl (-CH_3), methylene (-CH_2), methenyl (-CH=), formyl (-CH=O) or forminino (-CH=NH) groups. Folates act as cofactor carriers and donors of such 1-carbon units, and are especially important in the biosynthesis of adenine, guanine, thymidine and methionine, and in the metabolism of serine and histidine.

The importance of folate to humans is emphasised by deficiency-associated diseases such as megaloblastic anaemia, neural tube defects and cardiovascular disease.[1,2] While prokaryotic cells can synthesise folates, humans and other eukaryotes must obtain folates from the environment; hence their classification as vitamins.

The major route for cellular uptake of folate is the reduced folate carrier (RFC), which is therefore a pharmacological target for inhibiting cell growth. Mutations in this and other key proteins involved in folate metabolism may modify a cell's response to such drugs (e.g., methotrexate) and result in resistance.

Folates and folate metabolism

Folates are a family of molecules based on folic acid (pteroylglutamic acid, folacin). The name is derived from the Latin *folium* meaning leaf, and indicates the main dietary source as green-leafed vegetables.[3] It is also found in fruits, and is produced by the normal intestinal flora. The molecule is composed of a pterin linked to *para*-aminobenzoic acid (PABA), plus one or more glutamate residues (Figure 1).

Correspondence to: Dr G Bosson
Email: geoffrey.bosson@unn.ac.uk

A normal diet provides sufficient folates, in several chemical forms, to maintain health, growth and replication. Most dietary folates are polyglutamylated and must have the glutamate residues removed before the core molecule can be absorbed by the small intestine. Here, they are transported by the intestinal folate carrier (IFC-1) across the intestinal cells to the basolateral membrane, where they

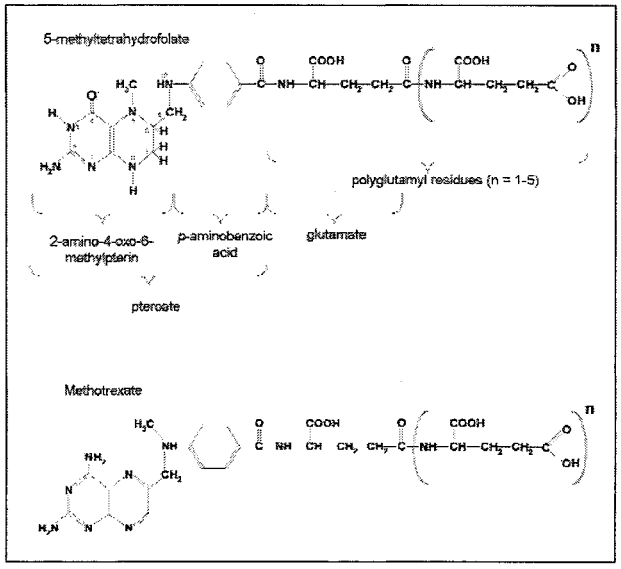


Fig. 1. Structural components of a typical folate molecule (5-methyltetrahydrofolate) and the structural similarity of the cytotoxic antifolate methotrexate.

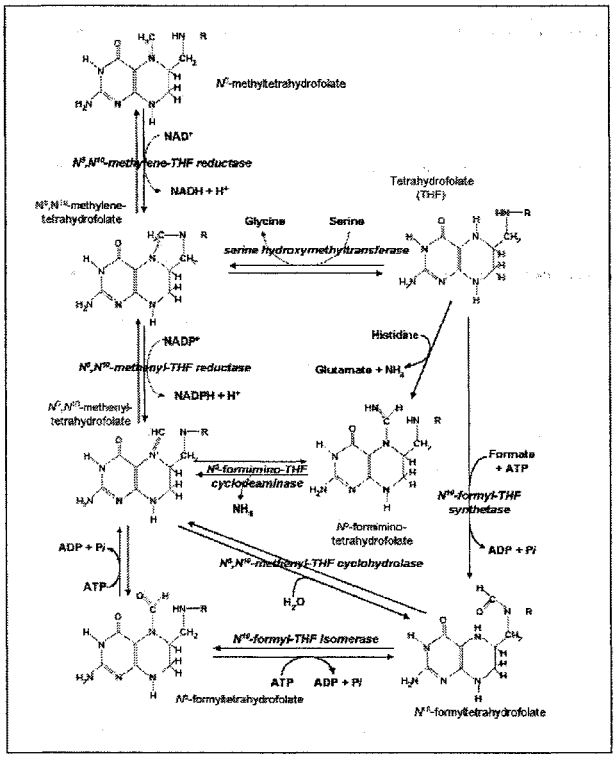


Fig. 2. Conversion of the 1-carbon residue carried by tetrahydrofolate through various oxidative states. For clarity, only the pterin residue is shown in each intermediary, the remaining PABA-glutamate residues are represented by R (modified from Voet *et al.*[3]).

diffuse along a concentration gradient into the submucosa.[4] It has also been suggested that an export mechanism at the basolateral membrane actively pumps folates from the cell and thus pulls folates from the intestinal lumen to replace those moving out of the cell.[5] In either mechanism, the folates delivered into the submucosa will diffuse into the portal circulation.

On first pass through the liver, hepatocytes remove and store the majority of folates until required. Any folates that reach the kidneys are filtered and then reabsorbed along the nephron.[4] The demands of the growing fetus for folates are met by transfer from the maternal circulation via a placental folate transporter (FOLT).

The natural circulating folates, N^5-methyltetrahydrofolate (5-methylTHF) and N^5-formyl-tetrahydrofolate (5-formylTHF) (Figure 2), are taken up by almost all cells in the body. At high concentration, folates can simply diffuse into the cell, but because folates are essentially hydrophilic molecules this does not provide sufficient intracellular folate for normal cell metabolism. To provide an adequate intracellular folate concentration, active transport mechanisms must be used that can work against a concentration gradient.

Folate receptors

Folate receptors (also known as folate binding proteins [FBP]) are glycosylphosphatidylinositol membrane-bound proteins with a high affinity for folic acid (K_M 1-10 nmol/L)[6] and 5-methylTHF (K_M 1-5 nmol/L).[7] Three FBP isoforms have been identified,[8] and the genes coding for two (FRa and FRb)

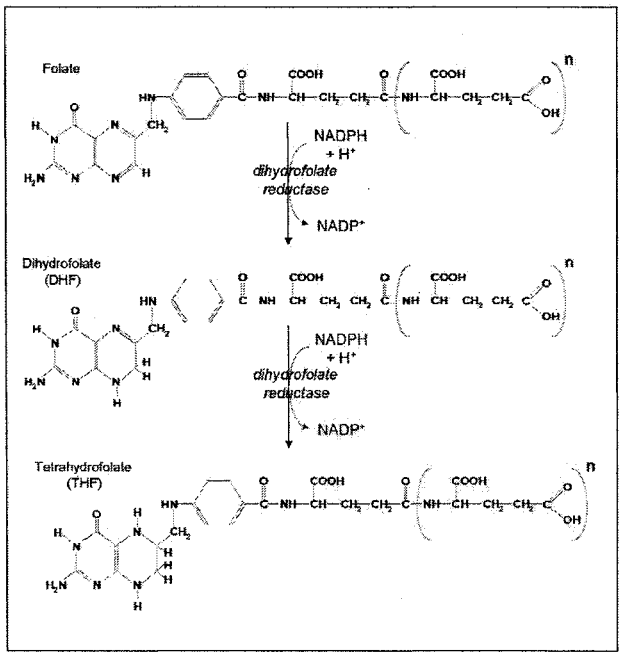


Fig. 3. Activation of folate to tetrahydrofolate (THF).

have been cloned.[9,10] When folic acid or 5-methylTHF binds to a receptor, the whole complex is internalised by endocytosis against the concentration gradient.

Reduced folate carrier

The folate receptors described above have a low affinity for reduced folates (K_M 10-300 nmol/L)[8] and these must therefore be brought into the cell using a different mechanism – RFC. This bi-directional anionic exchanger is the most important of the folate uptake mechanisms, with high affinity for reduced folates.[6]

The two systems may be coordinated to provide sufficient intracellular folates,[11] although some researchers believe they act independently.[12,13] In addition, there appears to be a third mechanism, able to recognise all folates with almost equal affinity (K_M 1-5 μmol/L), which operates at low pH.[14] Although its role is poorly understood, Rajgopal *et al.*[15] described the high efficiency of IFC-1 when operating at the low pH achieved by the acidic microenvironment at the luminal surface of the intestinal cells. The third mechanism may have a similar role, or prove to be one and the same thing. Mitochondria obtain folates from the cytoplasmic pool using an independent RFC in the mitochondrial wall.[16]

Once inside the cell, folylpolyglutamate synthetase (FPGS) adds glutamate residues to the C-terminal of the existing single glutamyl residue. This increases its size and prevents its loss from the cell via the folate export pumps.[14] Folic acid must be reduced by the enzyme dihydrofolate reductase (DHFR) to activate it in a two-step process to produce the fully active tetrahydrofolate (Figure 3). The same enzyme catalyses both steps but, because the preferred substrate is dihydrofolate, the first step is slower than the second.[17] Thus, this enzyme plays a pivotal role in the activation and regeneration of folates and makes it the target for drugs such as methotrexate (MTX).

One of the THF intermediates, N^5,N^{10}-methyleneTHF, catalysed by thymidylate synthase, donates its methylene

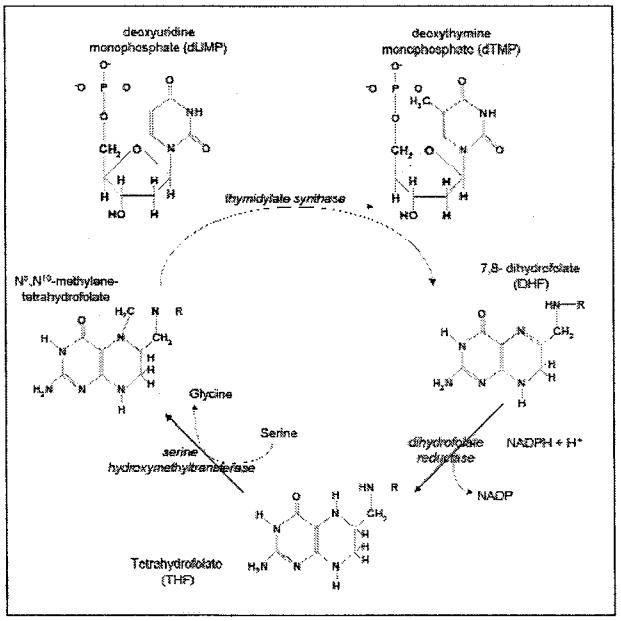


Fig. 4. Methylation of dUMP to dTMP and regeneration of dihydrofolate (DHF) to tetrahydrofolate (THF) by dihydrofolate reductase (DHFR). For clarity, only the pterin residue is shown in each intermediary.

group to deoxyuridine monophosphate (dUMP) to generate the cell's source of *de novo* deoxythymidylate monophosphate (dTMP). The remaining 7,8-dihydrofolate is recycled back to THF using DHFR (Figure 4).

The ability of living cells to synthesise *de novo* organic molecules is possible by adding carbon centres to each other or to existing biomolecules. It is in this role as carbon donor that tetrahydrofolate is so important, and is the result of its ability to hold the carbon in a variety of oxidative states, ranging from methyl (-CH_3) to methenyl (-CH=) groups. The 1-carbon residue can be modified from one oxidative state to another (Figure 2), which gives it great versatility in anabolic pathways.

Reduced folate carrier

The RFC is a 65 kDa transmembrane protein comprising 591 amino acid residues. In humans it exists in a highly glycosylated form with a mass of approximately 92 kDa.[18] While folate carriers are likely to be found in all higher animals, much of the data on RFC has been obtained from mouse and hamster models that show approximately 50% homology in amino acid sequence with the human protein.[19] A summary of characteristics between the three species is shown in Table 1.

Functionally, RFC is an anion exchanger that transfers hydrophilic folates across the cell membrane using the negatively charged glutamate residue of folate. Such a system requires positively charged amino acids and the arginine residues (at 133, 155 and 366 in murine cell lines) are highly conserved and likely candidates.[20] While the charge of these amino acids is important, Zhao *et al.*[20] showed that it is not the only factor when they substituted Arg-131 with

Table 1. Characteristics of RFC from the three species most frequently used in studies

Characteristic	Human	Hamster	Mouse
Size (protein only) (inc. glycosylated attachment)	65 kDa 85 kDa	58.6 kDa	58.2 kDa
pI	9.6	9.4	9.4
Number of amino acids	591 residues	518 residues	512 residues
Location of gene	Chromosome 21	Chromosome 1	Chromosome 10
Number of alternative spliced products	18 isoforms	2 isoforms	3 isoforms

histidine. Despite replacing one positively charged amino acid with another, activity was still lost.

While the protein structure of RFC in all three species has a conserved potential phosphorylation site for protein kinase-C at residue 23,[21] its use has yet to be uncovered. RFC lacks the structure typical of other transmembrane transport proteins that utilise ATP, which leaves the theory that intracellular phosphate is the driving force for folate uptake. High intracellular phosphate (PO_4^{2-}) is the anion that leaves the cell along a concentration gradient. To maintain neutrality, phosphate could be carried back into the cell but the protein has an affinity for reduced folates and extracellular phosphate concentration is low.[14] Once inside the cell, folate concentration is relatively low compared to phosphate, and very little leaves the cell via the RFC under competitive conditions.

RFC structure shown in Figure 5 is based on the Kyte-Doolitle hydropathy plot[22] and the Hopp and Wood hydrophilicity plot.[16] The putative transmembrane arrangement derived from this data is accepted by the major RFC research groups;[6, 14, 20, 23, 24] however, the complete three-dimensional (3-D) structure is yet to be elucidated. Figure 5 clearly shows the arrangement of 12 α-helical transmembrane domains made up of hydrophobic amino acids. The transmembrane domains (TMD) range in size from 17-25 residues and are highly conserved between the three species studied,[25] especially the nine tryptophan residues and four of the six cysteine residues.[25]

Hydrophilic amino acids make up the four small intracellular loops, six extracellular loops, C-terminal, *N*-terminal regions and the large intracellular loop formed halfway along the sequence. Both C- and *N*-terminals of normal RFC are shown lying in the cytoplasm. This was confirmed experimentally by the transfection of a known epitope construct into the C-terminal of a K562 cell line and subsequent localisation of the translated product by immunofluorescence staining for the inserted epitope.[26] The fluorescent label was bound to the cell membrane, but only after the cell was made permeable with Triton X-100.[26] Similar studies have been performed to confirm the position of other domains in relation to the cell membrane.[27]

Human RFC is larger than rodent RFC, due to a 73-79 amino acid C-terminal extension.[25] The terminal regions are not thought to play a role in the protein's function but, like many proteins, they contain motifs recognised for membrane localisation and intracellular trafficking. The RXR

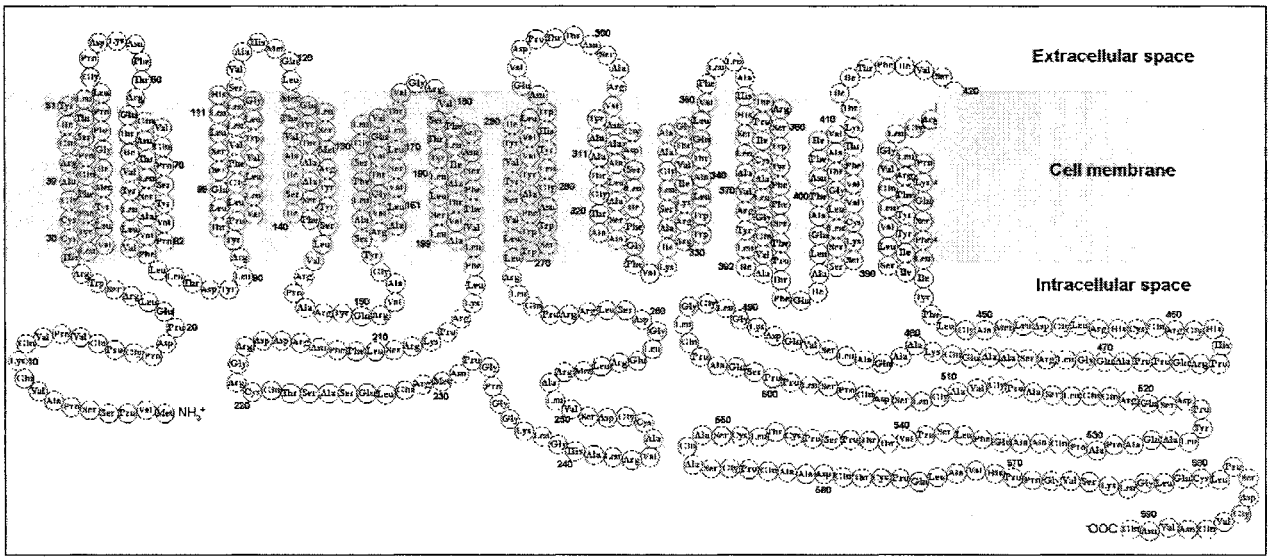


Fig. 5. Amino acid sequence and probable secondary structure producing 12 transmembrane domains of the RFC (modified from Drori *et al.*[6] and Zhao *et al.*[20]).

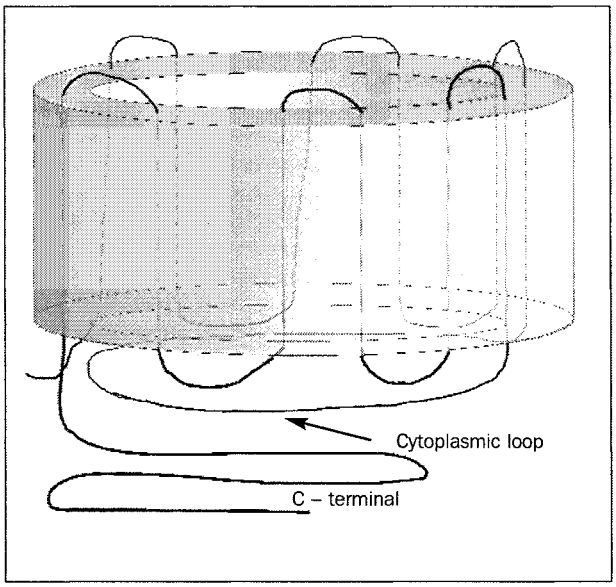


Fig. 6. Schematic to show how the large intracellular loop could play a role in anchoring the transmembrane domains into the expected cylindrical conduit in the cell membrane.

(Arg-X-Arg) motif at residue 460 causes temporary retention of RFC at the endoplasmic reticulum.[27] This sequence lies in the C-terminal and any mutations affecting or causing loss of the C-terminus will influence the time it takes for the protein to pass along the secretory pathway and embed itself in the cell membrane. This not only hinders correct positioning of the protein but also blocks any further protein synthesis. The C-terminal contains a di-leucine motif at residues 498 and 499 that is also likely to play a role in protein processing and localisation.[27]

The role of chaperones during the synthesis and transport of the protein has yet to be investigated but are likely to prove important – as they are for other proteins – in forming the appropriate tertiary structure. RFC maintains a stable tertiary structure using, among other bonds, charged pair attraction between negatively charged aspartate residues at 88 and 453 and positively charged arginine at 133 (TMD 4).[23]

Liu *et al.*[23] noted a significant drop in activity (V_{max}) when Asp88 was replaced with a neutral valine. However, when substituted with another negatively charged amino acid (glutamic acid), activity only dropped slightly. Similar replacement of the positive Arg133 by a neutral amino acid (leucine) caused loss of activity. It is interesting to note that when both residues 133 and 88 lose their charge the tertiary structure can still be formed and activity is maintained. This latter finding suggests that these residues are involved in stabilisation of the protein structure and are not required for direct recognition of folate or transport.

The size of the large intracellular loop also appears to be crucial to development of the correct RFC conformation.[27,28] This suggests that it could have a role in anchoring the TMDs into the cylindrical conformation expected of a transmembrane carrier (Figure 6).

While the concept is worthy of further study, recent work by Liu *et al.*[28] indicates that it is unlikely. They deleted sections of the loop and then replaced them with similar sections from the loop of the thiamine transporter (ThTr1) to form chimaeric constructs. With the exception of a short sequence of 11 amino acids (K_{204}RPKRSLFFNR$_{214}$), the majority of residues in the loop were not conserved. However, deletion of the whole or substantial parts of the loop significantly reduced RFC activity. Once passed the first 11 residues of the loop (the importance of which is unknown), the actual sequence of amino acids had minimal effect. These findings support the theory of Sadlish *et al.*[27] that only six TDMs can be inserted into the cell membrane at one time, as the loop keeps them far enough apart for two insertions but maintains the proximity of each.

Glycosylation on residue 58, asparagine, accounts for the significant difference in molecular weight described by Matherly and Angeles.[29] In the presence of tunicamycin to block glycosylation, or following elimination of the

N-glycosylation site by site-directed mutagenesis, RFC lost some of its activity.[30] The most likely explanation is that, as it migrates along the endoplasmic reticulum (ER), the usual glycosylation of proteins increases the viscosity of the ER lumen, causing longer retention and ultimately reduces the amount arriving at the cell membrane.[31] However, Wong *et al.*[30] concluded that glycosylation of RFC played no role in its function or its ability to migrate to the cell membrane and establish itself as a transmembrane structure.

RFC-1 gene

Once the protein structure was identified, it was relatively easy to generate complementary DNA (cDNA) for the protein and find a match in the human genome built from contig alignment. A sequence on chromosome 21 at SLC19A1 was identified and confirmed using fluorescence *in situ* hybridisation (FISH). The gene coding for the human RFC protein is located on the long arm of chromosome 21, specifically the 21q22.2-q22.3 region.[21,32]

The structure of the gene has been studied in mouse, hamster and human cell lines and is referred to as RFC-1. Tolner *et al.*[33] studied cDNA from 16 genomic clones and obtained two overlapping sequences covering the RFC-1 gene. The first strand, labelled λhRFC1-1, is a 19 kb segment containing exons 1, 2, 3 and 4. The second strand, λh-RFC1-2, is slightly shorter at 17 kb and contains exons 5 and 6. The 5′ end of λhRFC1-2 overlaps 1.5 kb of the 3′ end of λhRFC1-1 and thus together they cover 34.5 kb. The full-length cDNA was confirmed as the RFC-1 gene by transfection into a methotrexate-resistant cell line. These cells have become resistant to this folate analogue by producing a mutation in the protein, and no longer take up MTX. When the proposed gene is transfected into such cells and they start to transport MTX across the cell membrane, this is strong evidence that the genetic material introduced codes for RFC.

The exonic sequences have been submitted to both the National Center for Biotechnology Information (NCBI) and European Molecular Biology Laboratory (EMBL) databases, using the accession numbers shown in Table 2. Following rapid amplification of cDNA ends (RACE) and gene walking studies, the intronic regions between the exons have been defined and show several intron/exon/intron splice sites that comform to the GT-AC rule. The boundaries and exonic sequences are mainly conserved across the species, while the introns can vary greatly in size and sequence.

A shorter cDNA sequence was submitted to GenBank (accession number U19870) by Matherly's group,[25] and contained a 1776 bp open reading frame (ORF) flanked by a 98 bp 5′-UTR and 864 bp 3′-UTR. The cDNA ORF from either source predicts a protein of 64,873 Da, which is in the range for the deglycoslyated protein obtained experimentally and demonstrated by Western blot.[30]

As can be seen from Table 2, Tolner *et al.*[33] described three alternative untranslated sequences in exon 1, labelled 1a, 1b and 1c. Each ends with the AG sequence that provides the opportunity for alternative splicing and produces three RFC-1 variants: variant I contains exon 1a plus exons 2-6; variant II contains exon 1b plus exons 2-6; and variant III contains exon 1c plus exon 2 and part of 3.

Variants I and II are active proteins, but the truncated RFC from variant III is inactive. A truncated variant was noted in

Table 2. Accession numbers for RFC-1 exons on the NCBI Entrez (http://www.ncbi.nlm.nih.gov/Entrez/) and EMBL (http://www.ebi.ac.uk/embl/index.html) databases

Exon	Accession No.	Size (bp)
1a, 1b and 1c	U92868	3772
2	U92869	250
3	U92870	772
4	U92871	214
5	U92872	151
6	U92873	1451

hamster cell lines in 1994[34] and the suggestion that it was the result of alternative splicing first postulated. The 5′-end untranslated sequence contains promoters that have been studied using inserted downstream indicator genes (e.g., luciferase). Using this technique, it was shown that variant II produces the most efficient transcription.

Using different cell lines, Gong *et al.*[35] showed that the alternative splicing appeared to be tissue-specific and introduced a fourth alternative of exon 1. This new exon, 1d, was only found in fetal liver cells. Exons 1a and 1d are separated from their neighbours by large intronic regions and appear to have independent promoters. With only 21 bases between exons 1b and 1c, they are likely to share the same promotor upstream of exon 1b.

The tissue-specific differences recently highlighted by Whetstine *et al.*[36] also use the alternative splicing theory of 5′-UTRs. The seven possible non-coding exons contain either promoters or promotor enhancers that result in 18 unique splice variants; however, the role of each is yet to be completely understood.

While Tolner *et al.*[33] confirmed that variant III causes a truncated protein, a paper by Zhang *et al.*[18] in the same year proposed more subtle alterations in post-transcriptional processing without altering the coding sequence of the protein. To confuse exon identification further, Zhang *et al.*[18] used a different numbering system for the exons, whereby exon 3 contained the Kozak start sequence (ATG) and corresponds to exon 2 in Tolner's model.[33,37]

Although there is increased RFC-1 expression in response to very low folate concentration,[14] little is known about the regulation of RFC-1 expression. When 5-formyl folate is reintroduced to 'starved' cells they revert to normal transcription rates, suggesting that a feedback mechanism between intracellular folate level and RFC transcription must exist.[38] Gong *et al.*[35] measured messenger RNA (mRNA) levels throughout the normal cell cycle and noted that transcription is linked to cellular requirements for folate, reaching a peak during the G_1/S phase. In fact, they showed that the variable transcripts show a chronological pattern during the cell cycle, with exon 1c being expressed earlier than 1b and 1d. The significance of this is unknown.

In a series of experiments to determine the promoters for the RFC-1 gene, Zhang *et al.*[18] studied a 342 bp intronic region between exon 2 and 3 (1b and 2 in Tolner's model), which they labelled Pro32, and a 996 bp region upstream of exon 1 (1a in Tolner's model) was labelled Pro43. These regions were later reclassified as hRFC-A and hRFC-B, respectively.[39] The presence of two promoter regions for the

RFC-1 gene indicates the importance of the RFC protein and provides a mechanism to enhance transcription to a higher level when need for the protein is greatest.

Using the luciferase gene as an indicator sequence, these two promoter regions were studied to determine the exact mode of control. Sequential loss of 5′ upstream regions showed that promoters in Pro32 reside between -501 and -455 (where +1 is the translational start site). This 47 bp sequence is GC-rich and contains a CRE/AP-1 element at -485 to -471. The bZip DNA binding proteins (e.g., CREB-1, ATF-1 and c-Jun) bind to CRE/AP-1 elements and cause transcription as part of the cAMP secondary messenger system. An additional AP binding site is present in a recently identified variable nucleotide tandem repeat (VNTR), which increases promoter activity of the 5′-UTR by 63%.[40]

The more efficient Pro43 region that lies between -1088 and -1043 (46 bp) is even more GC-rich and, while there is no TATA box, the high GC content does provide a GC box (CCCGCCC) between -1081 to -1078. Whetstine *et al.*[39] also suggest that there is a second GC box between -1077 and -1071. GC boxes are known to bind Sp1-activating factors and Sp3-antagonistic factors. With opposing actions, Sp1 and Sp3 may be involved in fine control of gene expression and splice variant transcription.

The intronic sequence of the 5′-untranslated region (UTR) may also contain an inhibitory site between -2338 and -1935. This was indicated during the progressive deletion of regions to find the promoters, when increased expression was noted on removal of this particular section.[39]

It is interesting to note that the *p*53 gene may have a role in controlling RFC-1 expression. p53 protein is expressed when a defect in DNA (usually occuring during replication) is detected and inhibits cell replication until either the error is repaired or apoptosis is induced.[41] In this situation, it makes sense to decrease folate uptake and starve the cell of components necessary for replication. This is achieved, to some degree, by p53 suppression of RFC-1 expression.[42] p53 acts on the hRFC-B promoter rather than hRFC-A.[24]

While much effort has been devoted to the 5′-UTR, it appears that exon 6 at the other end may be just as variable. This would explain some of the differences seen between species for the cytoplasmic C-terminal.[43] Brigle *et al.*[44] reported similar C-terminal variation in murine RFC-1 that may play a role in controlling folate uptake. Restriction enzyme analysis of the murine gene indicates that the alleles are heterozygous.[44] The conclusion, therefore, is that if the wild-type allele is present in the cell lines studied, there must be some silencing of the normal allele. The proposed mechanism for this effect is DNA methylation, known to occur in suppression of other genes and also likely to occur in humans.[45]

Acute lymphoblastic leukaemia

Leukaemias are a group of diseases resulting from the uncontrolled proliferation of leucocytes. Broadly, they are classified as either acute or chronic, and subgroups including acute lymphoblastic leukaemia (ALL).[46] ALL originates in the bone marrow and by the time the clinical symptoms are evident it has replaced most of the normal marrow and metastasised.

Presentation of ALL is essentially the same in adults and children at relapse or original diagnosis. Pallor, malaise and weakness due to anaemia are the principal symptoms, but bruising, epistaxis and petechiea indicate associated thrombocytopenia. Both are the result of the normal marrow being replaced by malignant tissue, which also leads to increased susceptibility to infection as normal leucocyte production is disordered.

ALL is diagnosed twice as often in children than adults, but the former show a better response to treatment, with an 80% cure rate possible.[47] This success rate can be improved by initiating a more rigorous treatment regimen from the start,[48] but this requires tighter classification of the malignant cell line, particularly in relation to specific gene mutations.

Although some of the gene mutations leading to ALL can be identified by cytogenetics (e.g., translocations that lead to inactivation of suppressor genes – the Philadelphia chromosome being the most common), there are few specific morphological markers and classification is often based on immunological identification using cluster of differentiation (CD) markers.[49] Reverse transcription polymerase chain reaction (RT-PCR), FISH and other techniques available to the molecular biologist are increasingly important in identifying genes associated with poor prognosis (e.g., multiple drug resistance gene).[50]

There are three phases in the treatment of ALL: induction of remission (returns blood profile to normal), consolidation (keeps blood profile normal) and continuation. The final phase is rarely required for more than three years, but attempts to reduce this sadly correlate with a higher relapse rate.[50]

Relapse is more common in adults and is likely to be due to the higher frequency of genetic abnormalities in this group. Relapse in both children and adults can result from the lower drug concentrations in some body compartments (e.g., cerebrospinal fluid, testes) that are sanctuaries for the malignant cells.[49] Exposure to drug levels below the minimum effective dose gives the cells a chance to evolve and develop drug resistance. Proliferation of these rogue cells leads to relapse and a poor prognosis.

Methotrexate

Methotrexate (4-NH_2,*N*-10 methyl analogue of folic acid) is an effective and widely used cytotoxic agent. Rots *et al.*[47] describe how it has been used with success in the treatment of ALL since the 1940s. Similarity in structure between MTX and the folates (Figure 1) allows it to compete with and inhibit the key enzymes DHFR and TS. This inhibits purine and thymidine synthesis and regeneration of THF, thereby reducing DNA replication and cell proliferation.[51] The ratio of dUMP to dTMP is increased, making misincorporation during replication more likely, leading to fragmentation of the DNA.[47]

T-lineage ALL usually requires higher doses of MTX than do precursor B-ALL malignancies.[47] In high-dose therapy, used in some cases of MTX resistance, extracellular MTX concentrations >20 μmol/L are achieved regularly, and are high enough to facilitate simple diffusion across the cell membrane.[52] When lower routine doses of MTX are used, effective intracellular concentrations can only be achieved by using the RFC transport mechanism.

The anion exchange mechanism recognises the negatively

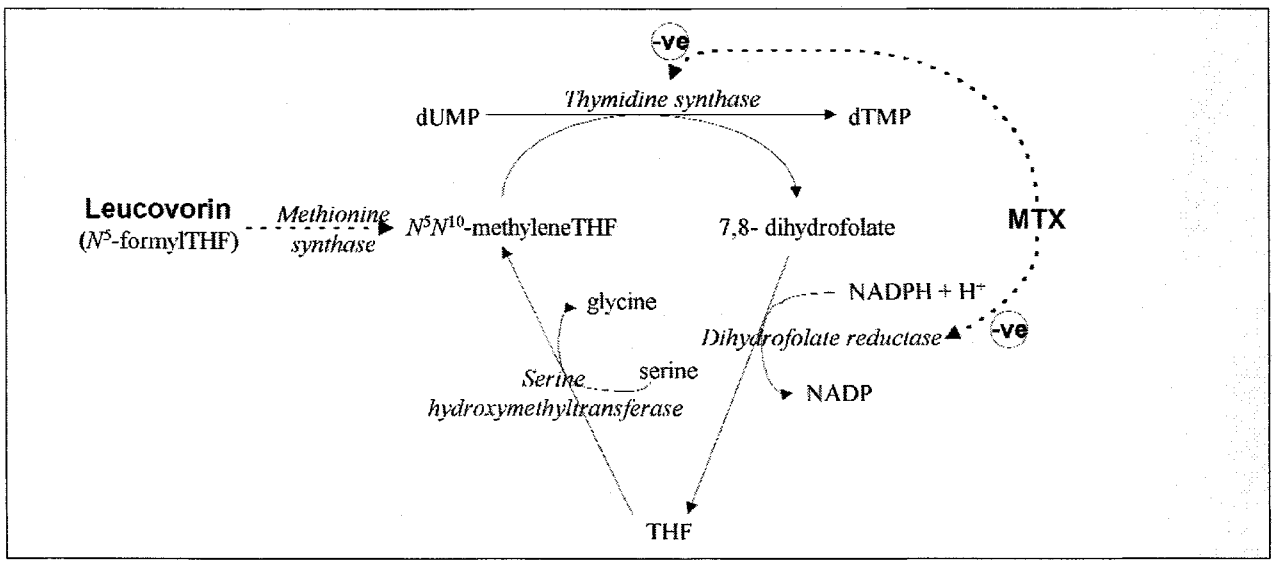


Fig. 7. Schematic of the reactions shown in Figure 4 to indicate where MTX and leucovorin are involved.

charged glutamate residue, which is also used to transport true folates. Once inside the cell, it must be polyglutamylated by folylpolyglutamate synthetase to prevent diffusion back out of the cell or active excretion via the multidrug resistance protein. Retention in the cell is proportional to the number of glutamyl residues.[53] MTX fits into the cleft of DHFR, but because the pteridine ring is inverted, it forms tighter bonds with the active site.[54] This means that the drug binds to the enzyme for longer, making it more effective. Another advantage of MTX polyglutamylation is that it directly inhibits TS.[53]

Mature cells have little demand for THF and are affected only minimally by MTX, but cells undergoing rapid replication (e.g., cells of the normal bone marrow and gastrointestinal mucosa, as well as cancer cells) take up the drug when their demand is highest, effectively poisoning themselves.[55] Cancer cells with hyperploidy for chromosome 21 express extra copies of the RFC-1 gene and have larger amounts of RFC in the cell wall, which makes them more sensitive to MTX.[56]

The most effective MTX regimen is high exposure over a long period of time, but the side effects are too severe. Slightly higher doses of MTX can be tolerated if leucovorin (N^5-formylTHF) is given to support the normal cells. This works because normal cells contain the enzyme methionine synthase, which converts N^5-formylTHF to N^5N^{10}-methyleneTHF, thus by-passing the need for DHFR to generate THF (Figure 7).

MTX's weak negative charge, which hinders its diffusion across cell membranes, also reduces its distribution into the CNS and other lipophilic tissues. Any cells in these sanctuaries will not be exposed to high enough concentrations of MTX to be effective, but they could be sufficiently high for the cells to develop a resistance.[53]

Methotrexate resistance

Natural selection in living things is based on the fact that mutations (which occur all the time) in a cell's genome may give it a survival advantage in a modified environment. Thus, if cells are exposed to MTX levels insufficient to be lethal, they will have a chance to mutate and replicate. If the mutation somehow decreases the intracellular concentration of active polyglutamylated MTX, then they will be able to tolerate higher doses in the future and survive standard treatment regimens (i.e., they will have developed a resistance).

The four possible mechanisms for resistance[8,47,53,57-59] are:

- decreased membrane transport into the cell by down-regulation of the RFC-1 gene, gross modification of the protein structure so that it is cleared from the membrane faster, impaired transport function, and decreased affinity of the RFC protein for MTX;
- altered DHFR response to MTX by decreased affinity for MTX, increasing true folates to overcome competitive inhibition by MTX,[60] and increasing the amount of DHFR;[59]
- MTX rendered less effective by reducing folylpolyglutamate synthetase activity; and,
- increased efflux of MTX from the cell by increasing folylpolyglutamate hydrolase activity,[47] and increasing activity of the multiple drug resistance protein.

While the final three mechanisms provide resistance once the drug is inside the cell, the first option – down-regulation of the RFC-1 gene – aims to prevent or reduce the amount of drug to which the intracellular systems are exposed, and is considered the first line of defence.

Down-regulation of the RFC-1 gene

The simplest and most responsive way to reduce the amount of RFC in the cell membrane is to suppress the gene coding for the protein.[61] MTX^{R}ZR-75-1 and K500E are MTX-resistant cell lines that have mRNA levels lower than those found in the corresponding wild-type cells.[21,62] One way to achieve

this is to simply down-regulate expression of the RFC-1 gene, possibly by a mutation in one of the upstream promoter regions, or a chromosomal translocation that produces transcriptional silencing.[62]

This suppression could be specific to particular spliced mRNA arrangements; for example, exon 1c was found to be the most frequently suppressed in studies by Gong *et al.*[35] The work by Ding *et al.*[41] into the control of RFC-1 expression by the *p*53 gene opens up the possibility that MTX resistance can be achieved indirectly by a mutation in the *p*53 gene.

Gorlick *et al.*[63] report decreased expression in a series of relapsed ALL cases, indicating that this mechanism is used *in vivo* and not just in 'forced' resistant *in vitro* cell lines. Measurement of RFC-1 mRNA is not just a research tool, it could also be used to determine those patients who will not respond to standard-dose MTX-based treatment.

Gross modification of the RFC protein structure

In resistant cell lines where the mRNA is normal, or even increased,[38] the explanation must be a modified protein. These modifications can be small (e.g., one amino acid) or gross changes in the structure as a whole and have tended to cluster into specific regions of the protein.[64]

A 7 bp deletion at the intron/exon 6 splice site (i.e., nucleotides 1152 to 1158) in the murine MtxRIIOua^{R}2-4 cell line causes the loss of two residues and a frame shift.[57] The amino acid sequence resulting from the frame shift has no homology with the original, and a premature stop codon is introduced. The resulting truncated protein with a modified C-terminal is less stable and, while normal RFC in the membrane has a turnover of approximately 50% per day,[65] the defective protein is cleared from the membrane faster than normal RFC.[57,65]

In humans, similar truncating mutations have been demonstrated in MTX-resistant sublines of MOLT-3[58] and CCRF-CEM.[65] The MOLT-3/MTX$_{10,000}$ cell line contains two truncating mutations.[58] One is the result of a G to A mutation at position 74 that changes the TGG codon (tryptophan) at position 25 to a stop codon, TAG. The second is a C to T transition involving codon 40, CAG (glutamine) to TAG. When both mutations occur together, one in each allele, the resistance is high (i.e., 10,000-fold). The CEM/Mtx-1 cells, studied by Wong *et al*,[65] have a frameshift that also produces an early stop codon. The four-base CATG segment inserted at position 191 causes a heterologous amino acid sequence, (similar to that described in the MtxRIIOua^{R}2-4 cells[57]) and an early stop codon at position 1176. The hydropathy plot of the translated mutant sequence no longer favours stable localisation in the cell membrane, and is therefore not active.

Not all truncated RFCs produce MTX resistance. A 987 base deletion in the CEM-7A leukaemic cell line results in the loss of 160 normally coded amino acids from the C-terminal and the normal stop codon, (Figure 8). The deletion of bases 1389 to 2376 allows the 5'-end sequence from base 2377 to be spliced to 1388, and produces an alternative ORF. This alternative sequence now codes for 58 new C-terminal residues before reaching a new stop codon in the previously untranslated 3'-region. The proposed structure of the resulting RFC is similar to that shown in Figure 5 but, because it lacks the last transmembrane domain, the truncated C-terminal end is left outside the cell membrane.[6]

The truncated protein alone is unable to transport MTX and correctly considered as non-functional by Matherly.[24]

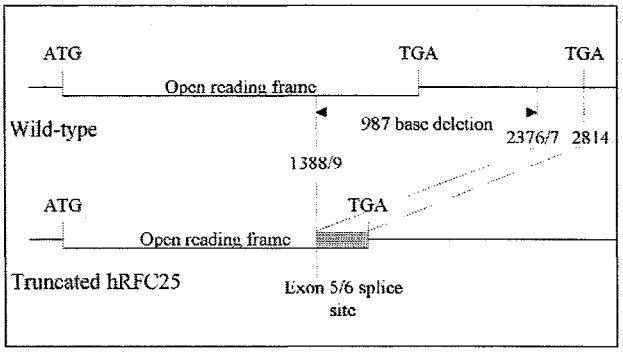


Fig. 8. Alternative splicing of cDNA from human leukaemic CEM-7A cells, producing a truncated RFC.[6]

However, Drori *et al.*,[6] in their discussion, suggest that over-expression of this 'alternatively spliced gene' and subsequent protein synthesis in heterozygotes may have modified the influx/efflux rates that render the cell MTX-sensitive. Brigle *et al.*[44] describe a similar 'alternatively spliced' variant in their studies of mouse cells, which lack several C-terminal end transmembrane domains.

Antibodies against RFC will not detect the truncated proteins described because the epitope sequence is missing in most cases. If immunoassay does detect a full-length translated protein then MTX resistance is likely to be the result of a single but critical residue alteration. These single amino acid alterations can affect the transport function of the protein, or its recognition and affinity for MTX/folates. A negative immunoassay does not exclusively indicate a truncated protein, because it is possible that a single amino acid substitution in the epitope will prevent antigen-antibody recognition.

Proline residues in the α-helical TMD are important features that the RFC should adopt in its tertiary structure. Thus, it is not suprising that the introduction of more proline, or substitution of existing proline, will produce kinks and significantly alter the protein structure. When this is done by site-directed mutagenesis, a loss of transport activity is noted, but affinity is unaffected.[65] This fits with the theory that the transmembrane domains are not involved directly in folate recognition.

Impaired transport function

A predicted conformational change in the RFC to transport folates and MTX requires several of the transmembrane domains to interact with intracellular domains.[66] Previously, the connecting intracellular loop sequences between the TMDs were thought to be just that, connectors, but when the primary structure was modified by the insertion of an immunogenic marker sequence, RFC activity was modified.[26] For example, insertion of this sequence into the loop between the 11th and 12th TMDs caused a total loss of activity. While the result indicates that the insertion of additional amino acids will push the TMDs sufficiently far apart to disrupt RFC function, such large insertions are not likely to occur *in vivo*.

When the charge on residue 45 is changed from negative glutamate to positive lysine, a decrease in the fluidity of the protein is thought to occur, rendering it non-functional. Evidence for this lies in the fact that chloride ions appear to

Table 3. Summary of mutations in the RFC-1 gene affecting the function of the RFC protein (i.e., change in V_{max} but not K_M)

Nucleotide affected	Amino acid residue affected	Species	Domain affected and effect on RFC	Ref
G133A	Glu45Lys	m	1st TMD. Makes the protein rigid	66
G137A	Ser46Asn	m	1st TMD. Allosteric restriction of RFC mobility	68
G175A	Arg27His	H	TMD1. Truncated at 225	64
C264G	Asp88Glu	H	2nd TMD. Slightly $\downarrow V_{max} \rightarrow K_M$	23
G388C	Ala130Pro	m	4th TMD. Proline causes a kink in the protein structure	65
G398A	Arg133His	H	4th TMD. Total loss of activity	23
G398T	Arg133Leu	H	4th TMD. Loss of ion pair with Asp 88 causing total loss of activity	23
CGC397-9GAG	Arg133Glu	H	4th TMD. Total loss of activity	23
C769A	Ser225stop	H	Truncated protein	64
A1033G	Gly345Arg	h	9th TMD. The introduction of the +ve charge of the arginine modifies the 2°/3° structure. Decreased half-life due to poor stability	57
G1118A	Arg373His	m	10th TMD. Loss of positively charged arginine reduces transport as much as 50-fold	67
AAG1210 – 1212CTG	Lys404Leu	m	11th TMD. Loss of positive charged lysine removes the inhibitory effect of Cl- on MTX uptake	69

Footnote
m = mouse, h = hamster, H = human
Nucleotide numbering is taken from the start codon of the open reading frame

Table 4. Summary of mutations in the RFC-1 gene affecting the binding affinity (i.e., V_{max} remains the same but K_M changes)

Nucleotide affected	Amino acid residue affected	Species	Domain affected and effect on RFC	Ref
G131A	Gly44Glu	m	1st TMD	20
G227A	Glu45Lys	H	1st TMD. Modifies sensitivity to Cl-	45
G133A	Glu45Gln	m	1st TMD. $\uparrow K_M$ for MTX, $\downarrow K_M$ for natural folates	70
GAA133 – 135CGT	Glu45Arg	m	1st TMD. $\rightarrow K_M$ for MTX, $\downarrow K_M$ for natural folates, significantly for folic acid	70
GAA133 – 135GAC	Glu45Asp	m	1st TMD. $\uparrow K_M$ for all substrates, including MTX	70
GAA133 – 135CTC	Glu45Leu	m	1st TMD. $\uparrow K_M$ for all substrates, including MTX	70
GAA133 – 135TGG	Glu45Try	m	1st TMD. $\uparrow K_M$ for reduced folates, but $\downarrow K_M$ for folic acid	70
G137A	Ser46Asn	m	1st TMD. $\uparrow K_M$ for MTX, $\downarrow K_M$ for natural folates 68	
A142T	Ile48Phe	m	1st TMD. $\rightarrow K_M$ for MTX, $\downarrow K_M$ for folic acid. $\uparrow K_M$ for $N^{5,10}$dideazatetrahydafolate	71
T313G	Trp105Gly	m	3rd TMD. $\downarrow K_M$ for folic acid	71
G890A	Ser297Asn	m	External loop between TMD 7-8. $\uparrow K_M$ 72	
C926T	Ser309Phe	m	8th TMD	73

Footnote
m = mouse, H = human
Nucleotide numbering is taken from the start codon of the open reading frame

neutralise the lysine and return function.[66] Larger anions, (e.g., sulphate, ATP) are too large to reach the lysine and have no neutralising capacity. The ability of RFC to function when this amino acid is neutralised indicates that the negatively charged glutamate at position 45 is not important to the binding of folates or to transport activity.

Forced mutations of Arg133 and Asp88[23] provide further evidence that TMDs interact with each other to produce a functional carrier. The two oppositely charged amino acids form an ion-pair. When both are mutated to neutral amino acids they can still form an association and have minimal effect on activity, but if just one of the pair is changed then it prevents the tertiary structure being formed and results in a total loss of activity.

The conclusion was complicated somewhat by the finding that substitution of the negatively charged Asp88 with Glu,

another negatively charged amino acid, produced a drop in activity. This suggests that the additional methylene group of glutamate is sufficiently large to cause a slight structural change, and that the conformation of the protein at this point is not dependent solely on charge interaction.[23]

Similar experiments by Sadlish *et al.*[67] on Arg373 in murine cells indicated the charge of this residue was just as important. The accepted topology of RFC puts this polar amino acid in TMD10, a non-polar region, and therefore suggests that this residue could play a role in temporarily holding the polar folates on their way through the RFC. A summary of these mutations is shown in Table 3.

Modified binding

The most effective way for a cell to obtain sufficient reduced folate to sustain a high replication rate, yet exclude MTX from it, is to make the RFC selective. This can be achieved by substituting amino acid residues responsible for binding recognition or affinity, and is demonstrated experimentally by changes in K_M. A summary of mutations that modify K_M are shown in Table 4.

Amino acids 45-48 appear to form an aqueous pocket in the first TDM that would be important for the binding of folates and MTX. It is not surprising, therefore, that mutations in the first TMD have an effect on affinity, some directly and others indirectly.

Results of targeted mutation studies performed by Zhao *et al.*[70] (Table 5) show that substitution of the negatively charged glutamic acid at position 45 with the negatively charged amino acid aspartic acid significantly decreases affinity for all four compounds. This indicates that it is not the negative charge at position 45 that is essential for RFC function. If substituted by a hydrophobic amino acid (i.e., leucine or tryptophan) then affinity is also decreased, but substitution with hydrophilic glutamine provides decreased affinity for MTX while increasing affinity for folates. The benefit of a hydrophilic amino acid gives support to the aqueous pocket theory.

The Glu45Arg substitution has little effect on the affinity for MTX, and relies on the increased affinity for reduced folates and folic acid to provide the resistance (i.e., by selecting their uptake in preference to MTX). The sensitivity of RFC to chloride ions in the Glu45Lys substitution is thought to result from an allosteric effect caused by small anions on a remote domain of the protein.[66]

Experiments by Sharina *et al.*,[69] who studied the chloride effect on other positively charged residues, showed that the substitution of lysine at residue 404 of murine RFC resulted in chloride losing the competitive inhibitory effect seen in wild-type RFC. The conclusion was that this residue could be a site for control of RFC activity by chloride ions.

Further evidence that the first TMD is involved in the binding function of RFC is provided by Ser46Asp[68] and Ile48Phe[71] mutations, which show similar transport selectivity. Work by both Zhao *et al.*[66] and Tse *et al.*[71] suggests that the third TMD may also play a part in forming this hydrophilic pocket. The increased affinity for folic acid seen in the Trp105Gly substitution provides resistance to another antifolate, $N^{5,10}$dideazatetrahydrofolate (DDATHF). This is achieved by the selective uptake of folic acid to increase the intracellular concentration, which saturates PFGS and prevents DDATHF taking its active polyglutamylated form.[71]

Not all mutations identified so far affect RFC activity. Some will be silent mutations (where the triplet code is changed, but the same amino acid is coded for) or the substitution affects a non-crucial amino acid. Table 6 provides a summary of mutations with no known affect. While these may not be useful to those studying MTX resistance, they may help to elucidate the 3-D structure of RFC.

Table 5. Effect on binding affinity of MTX, N^5– formylTHF, N^5– methylTHF and folic acid when glutamate at residue 45 is substituted (modified from Zhao *et al.*[70])

	K_M (μmol/L)			
Mutation	MTX	N^5– formylTHF	N^5– methylTHF	Folic acid
Wild type	7.0	5.6	1.5	260
Glu45Gln	15.8	1.6	0.6	42.0
Glu45Arg	8.0	1.2	1.4	27.0
Glu45Asp	25.0	50.0	4.3	400.0
Glu45Leu	13.0	11.4	6.5	117.0
Glu45Try		163.0	15.0	130.0

Table 6. Summary of mutations in the RFC-1 gene with no significant effect on folate or MTX transport

Nucleotide affected	Amino acid residue affected	Species	Domain affected and effect on RFC	Ref
G80A	Arg27His	H	1st TMD 1,	71
AAC171-4CAG	Asn58Gln	H	Extracellular loop connecting TMD 1-2. Removes the *N*-glycosylation site	30
A263T	Asp88Val	H	2nd TMD	23
CGC397-9AAG	Arg133Lys	H	4th TMD	23
C352T	Leu86Leu	H	Intracellular loop connecting TMD 2-3. Silent mutation	45
C696T	Pro232Pro	H	Silent mutation	74
C1242A	Ile414Ile	H	Silent mutation	74

Footnote
H = human
Nucleotide numbering is taken from the start codon of the open reading frame

Summary

Understanding of the biochemistry and molecular biology of cell metabolism means that new drugs can be designed to overcome resistance. RFC is the principal route for the uptake of reduced folates and cytotoxic antifolates into cells and is therefore one site at which resistance can be achieved.

Resistance to antifolate drugs, such as MTX, may be the result of altered RFC affinity, synthesis of a functionally inadequate carrier, or bypass of a transporter altogether by down-regulating RFC-1 expression and translation.

Trimetrexate is one of the new generation of antifolates produced using this information. It lacks the glutamate residue, leaving it non-polar and able to diffuse directly across the cell membrane (i.e., it bypasses any resistant mutation of RFC). Another strategy of therapeutic potential is the restoration of MTX sensitivity using transfection of intact mRFC.[76]

The majority of studies considered here have been performed on specific cell lines grown in the presence of different folates and/or antifolates. While these are 'forced' *in vitro* mutations, they do provide invaluable information on the structure and function of RFC. They are also believed to mimic *in vivo* resistance, and hopefully further studies will characterise clinically significant naturally occurring mutations.

Although we are still a long way from confirming the 3-D structure of the RFC, mutational studies indicate that the 12 TMDs form a cylindrical structure, which would form a channel in the cell membrane through which folates can be passed. The hydrophilic pocket formed by amino acids of TMD 1 and 3 at the extracellular surface of this channel is responsible for binding reduced folates by recognising and binding the pteridine domain. Here it waits for exchange with inorganic phosphate leaving the cell. To ensure it is transferred successfully through the channel, a series of key positively charged amino acids on the TMDs exposed to the channel are probably used.

Variety in the human genome is becoming apparent, with new single nucleotide polymorphisms (SNPs) revealed regularly. Important proteins, such as RFC, have highly conserved genes, which may go some way to explain why so few SNPs have yet to be assigned to this sequence.

As the proteomics age gathers pace, elucidation of protein structure and function will become commonplace. This will prove necessary to provide biomedical scientists, biotechnologists and clinicians with information to better understand disease, develop new tests, design new drugs and, ultimately, provide more effective treatment of disease.

References

1 Chango A, Emery-Fillon N, Potier de Courcy G *et al.* A polymoprhism (80G-A) in the reduced folate carrier gene and its association with folate status and homocysteinema. *Mol Genet Metab* 2000; **70**: 310-5.

2 Salooja N, Catto A, Carter A, Tudnesham EGD, Grant PJ. Methylene tetrahydrofolate reductase C677T genotype and stroke. *Clin Lab Haematol* 1998; **20**: 357-61.

3 Voet D, Voet J, Pratt CW. *Fundamentals of biochemistry* New York: John Wiley & Sons, 1999: 631-2, 693-721.

4 Wang Y, Zhao R, Russell RG, Goldman D. Localization of the murine reduced folate carrier as assessed by immunohistochemical analysis. *Biochim Biophys Acta* 2000; **1513**: 49-54.

5 Zeng H, Liu G, Rea PA, Kruh GD. Transport of amphipathic anions by human multidrug resistance protein 3. *Cancer Res* 2000; **60**: 4779-84.

6 Drori S, Sprecher H, Shemer G, Jansen G, Goldman ID, Assaraf YG. Characterization of a human alternatively spliced truncated reduced folate carrier increasing folate accumulation in parental leukemia cells. *Eur J Biochem* 2000; **267**: 690-702.

7 Trippett T, Schlemmer S, Elisseyeff Y *et al.* Defective transport as a mechanism of acquired resistance to methotrexate in patients with acute lymphocytic leukemia. *Blood* 1992; **80**: 1158-62.

8 Trippett TM, Bertino JR. Therapuetic strategies targeting proteins that regulate folate and reduced folate transport. *J Chemother* 1999; **11**: 3-10.

9 Brigle KE, Spinella MJ, Westin EH, Goldman ID. Increased expression and characterization of two distinct folate binding proteins in murine erythroleukemia cells. *Biochem Pharmacol* 1994; **47**: 337-45.

10 Wang X, Shen F, Freisheim JH, Gentry LE, Ratnam M. Differential stereospecificities and affinities of folate receptor isoforms for folate compounds and antifolates. *Biochem Pharmacol* 1992; **44**: 1898-1901.

11 McKusick VA. Solute carrier family 19, member 1: SLC19A1. *Medelian inheritance in Man.* Johns Hopkins University. <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=600424>

12 Westerhof GR, Jansen G, van Emmerik N *et al.* Membrane transport of natural folates and antifolate compounds in murine L1210 leukemia cells: role of carrier- and receptor-mediated transport systems. *Cancer Res* 1991; **51**: 5507-13.

13 Dixon KH, Mulligan T, Chung K-N, Elwood PC, Cowan KH. Effects of folate receptor expression following stable transfection into wild type and methotrexate transport-deficient ZR-75-1 human breast cancer cells *J Biol Chem* 1992; **267**: 24140-7.

14 Sierra EE, Goldman ID. Recent advances in the understanding of the mechanism of membrane transport of folates and antifolates. *Semin Oncol* 1999; **26** (Suppl 6): 11-23.

15 Rajgopal A, Sierra EE, Zhao R, Goldman ID. Expression of the reduced folate carrier SLC19A1 in IEC-6 cells result in two distinct transport activities. *Am J Physiol Cell Physiol* 2001; **281**: C1579-C1586.

16 Trippett TM, Garcia S, Manova K *et al.* Localization of a human reduced folate carrier protein in the mitochondrial as well as the cell membrane of leukemia cells. *Cancer Res* 2001; **61**: 1941-7.

17 Horton HR, Moran LA, Ochs RS, Rawn JD, Scrimgeour KG. *Principles of biochemistry* (3rd Ed). Englewood Cliffs, NJ: Neil Patterson Publishers (Prentice-Hall), 1993: 582.

18 Zhang L, Wong SC, Matherly LH. Transcript heterogeneity of the human reduced folate carrier results from the use of multiple promoters and variable splicing of alternative upstream exons. *Biochem J* 1998; **332**: 773-80.

19 Williams FMR, Flintoff WF. Isolation of a human cDNA that complements a mutant hamster cell defective in methotrexate uptake. *J Biol Chem* 1995; **270**: 2987-92.

20 Zhao R, Sharina IG, Goldmann ID. Pattern of mutations that results in loss of reduced folate carrier function under antifolate selective pressure augmented by chemical mutagenesis. *Mol Pharmacol* 1999; **56**: 68-76.

21 Moscow JA, Gong M, He R *et al.* Isolation of a gene encoding a human reduced folate carrier (RFC1) and analysis of its expression in transport-deficient, methotrexate-resistant human breast cancer cells. *Cancer Res* 1995; **55**: 3790-4.

22 Dixon KH, Lampher BC, Chiu J, Kelly K, Cowan KH. A novel cDNA restores reduced folate carrier activity and methotrexate sensitivity to transport deficient cells. *J Biol Chem* 1994; **269**: 17-20.

23 Liu XY, Matherly LH. Functional interaction between arginine-133 and aspartate-88 in the human reduced folate carrier: evidence for a charge-pair association. *Biochem J* 2001; **358**: 511-6.

24 Matherly LH. Molecular and cellular biology of the human reduced folate carrier. *Prog Nucleic Acid Res Mol Biol* 2001; **67**: 131-62.

25 Wong SC, Proefke SA, Bhushan A, Matherly LH. Isolation of human cDNAs that restore methotrexate sensitivity and reduced folate carrier activity in methotrexate transport-defective chinese hamster ovary cells. *J Biol Chem* 1995; **270**: 17468-75.

26 Ferguson PL, Flintoff WF. Topological and functional analysis of the human reduced folate carrier by hemagglutinin epitope insertion. *J Biol Chem* 1999; **274**: 16269-78.

27 Sadlish H, Williams FM, Flintoff WF. Cytoplasmic domains of the reduced folate carrier are essential for trafficking, but not function. *Biochem J* 2002; **364**: 777-86.

28 Liu XY, Witt TL, Matherly LH. Restoration of high-level transport activity by human reduced folate carrier/ThTr1 chimeric transporters: role of the transmembrane domain 6/7 linker region in reduced folate carrier protein. *Biochem J* (ePublished: BJ20020419).

29 Matherly LH, Angeles SH. Role of N-glycosylation in the structure and fucntion of the methotrexate membrane transporter from CCRF-CEM human lymphoblastic cells. *Biochem Pharmacol* 1994; **47**: 1094-8.

30 Wong SC, Zhang L, Proefke SA, Matherly LH. Effects of the loss of capacity for N-glycoslylation on the transport activity and cellular localization of the human reduced folate carrier. *Biochim Biophys Acta* 1998; **1375**: 6-12.

31 Nehls S, Snapp EL,Cole NB *et al.* Dynamics and retention of misfolded proteins in native ER membranes. *Nat Cell Biol* 2000; **2**: 288-95.

32 Lapenta V, Sossi V, Gosset P *et al.* Construction of a 2.5-Mb integrated physical and gene map of distal 21q22.3. *Genomics* 1998; **49**: 1-13.

33 Tolner B, Roy K, Sirotnak FM. Structural analysis of the human RFC-1 gene encoding a folate transporter reveals multiple promoters and alternatively spliced transcripts with 5′ end heterogeneity. *Gene* 1998; **211**: 331-41.

34 Williams FMR, Murray RC, Underhill M, Flintoff WF. Isolation of a hamster cDNA clone coding for a function involved in methotrexate uptake. *J Biol Chem* 1994; **269**: 5810-6.

35 Gong M, Cowan KH, Gudas J, Moscow JA. Isolation and characterization of genomic sequence involved in the regulation of the human reduced folate carrier gene (RFC1). *Gene* 1999; **233**: 21-31.

36 Whetstine JR, Flatley RM, Matherly LH. The human reduced folate carrier gene is ubiquitously and differentially expressed in normal human tissues: identification of seven non-coding exons and characterization of a novel promoter. *Biochem J* 2002; **367**: 629-40.

37 Tolner B, Roy K, Sirotnak F M. Organization, structure and alternate splicing of the murine RFC-1 gene encoding a folate transporter. *Gene* 1997; **189**: 1-7.

38 Jansen G, Westerhoff GR, Jarmuszewski MJA, Kathmann I, Rijksen G, Schornagel JH. Methotrexate transport in variant human CCRF-CEM leukemia cells with elevated levels of the reduced folate carrier. *J Biol Chem* 1990; **265**: 18272-7.

39 Whetstine JR, Matherly LH. The basal promoters for the human reduced folate carrier gene are regulated by a GC-box and a cAMP-response element/AP-1-like element. *J Biol Chem* 2001; **276**: 6350-8.

40 Whetstine JR, Witt TL, Matherly LH. The human reduced folate carrier gene is regulated by the AP2 and Sp1 transcription factor families and a functional 61 base pair polymorphism. *J Biol Chem* 2002 (ePublished: M208296200).

41 Ding BC, Whetstine JR, Witts TLM Schuetz JD, Matherly LH. Repression of human reduced folate carrier gene expression by wild-type p53. *J Biol Chem* 2001; **276**: 8713-9.

42 Lohrum MAE, Vousden KH. Regulation and function of the p53-related proteins: same family different rules. *Trends Cell Biol* 2000; **10**: 197-202.

43 Zhang L, Wong SC, Matherly LH. Structure and organization of the human reduced folate carrier gene. *Biochem Biophys Acta* 1998; **1442**: 389-93.

44 Brigle KE, Spinella MJ, Sierra EE, Goldman ID. Organization of the murine reduced folate carrier gene and identification of variant splice forms. *Biochem Biophys Acta* 1997; **1353**: 191-8.

45 Jansen G, Mauritz R, Drori S *et al.* A structurally altered human reduced folate carrier with increased folic acid transport mediates a novel mechanism of antifolate resistance. *J Biol Chem* 1998; **273**: 30189-98.

46 Hanson CA. Haematology. In: McLactchey KD, ed. *Clinical laboratory medicine*. London: Williams and Wilkins, 1994: 817-63, 939-69.

47 Rots MG, Pieters R, Kaspers GJL, Veerman AJP, Peters GJ, Jansen G. Classification of *ex vivo* methotrexate resistance in acute lymphoblastic and myeloid leukaemia. *Br J Haematol* 2000; **110**: 791-800.

48 Rots MG, Pieters R, Peters GJ *et al.* Methotrexate resistance in relapsed childhood acute lymphoblastic leukaemia. *Br J Haematol* 2000; **109**: 629-34.

49 Mauer AM. Acute lymphocytic leukemia. In: Beutler E, Lichtman M A, Coller BS, Kipps TJ, eds. *William's hematology* (5th Ed). New York: McGraw Hill, 1995: Chapter 105.

50 Pui C-H, Evans WE. Acute lymphoblastic leukaemia. *N Engl J Med* 1998; **339**: 605-15.

51 Belur LR, Boelk-Galvan D, Diers MD, McIvor RS, Zimmerman CL. Methotrexate accumulates to similar levels in animals transplanted with normal versus drug-resistant transgenic marrow. *Cancer Res* 2001; **61**: 1522-6.

52 Warren RD, Nichols AP, Bender RA. Membrane transport of methotrexate in human lymphoblastoid cells. *Cancer Res* 1978; **38**: 668-71.

53 Jolivet J, Cowan KH, Curt GA, Clendeninn NJ, Chabner BA. The pharmacology and clinical use of methotrexate. *N Engl J Med* 1983; **309**: 1094-104.

54 Blakley R L, Sorrentino B P. *In vitro* mutations in dihydrofolate reductase that confers resistance to methotrexate: Potential for clinical application. *Hum Mutat* 1998; **11**: 259-63.

55 Hoffbrand AV, Tripp E, Catovsky D, Das KC. Transport of methotrexate into normal haemopoietic cells and into leukaemic cells and its effects on DNA synthesis. *Br J Haematol* 1973; **25**: 497-511.

56 Masson E, Relling MV, Synold TW *et al.* Accumulation of methotrexate polyglutamates in lymphoblasts is a determinant of antileukemic effects *in vivo*, a rationale for high-dose methotrexate. *J Clin Invest* 1996; **97**: 73-80.

57 Sadlish H, Murray RC, Williams MR, Flintoff WF. Mutations in the reduced-folate carrier affect protein localization and stability. *Biochem J* 2000; **346**: 509-18.

58 Gong M, Yess J, Connolly T *et al.* Molecular mechanism of antifolate transporter-deficiency in a methotrexate-resistant MOLT-3 human leukemia cell line. *Blood* 1997; **89**: 2494-9.

59 Ohnuma T, Lo RJ, Scanlon KJ *et al.* Evolution of methotrexate resistance of human acute lymphoblastic leukemia cell *in vitro*. *Cancer Res* 1985; **45**: 1815-22.

60 Peters GJ, Jansen G. Folate homeostasis and antiproliferative activity of folates and antifolates. *Nutrition* 2001; **17**: 737-8.

61 Zhang L, Taub JW, Williamson M *et al.* Reduced folate carrier gene expression in childhood acute lymphoblastic leukemia: relationship to immunophenotype and ploidy. *Clin Cancer Res* 1998; **4**: 2169-77.

62 Ding BC, Witt TL, Hukku B, Heng H, Zang L, Matherly LH. Association of deletions and translocations of the reduced folate carrier gene with profound loss of gene expression in methotrexate-resistant K562 human erythroleukemia. *Biochem Pharmacol* 2001; **61**: 665-75.

63 Gorlick R, Goker E, Trippett T *et al.* Defective transport is a common mechanism of acquired methotrexate resistance in acute lymphocytic leukemia and is associated with decreased reduced folate carrier expression. *Blood* 1997; **89**: 1013-8.

64 Rothem L, Ifergan I, Kaufman Y, Priest DG, Jansen G, Assaraf YG. Resistance to multiple novel antifolates is mediated via defective drug transport resulting from clustered mutations in the reduced folate carrier gene in human leukaemia cell lines. *Biochem J* 2002; **367**: 741-50.

65 Wong SC, Zhang L, Witt TK, Proefke SA, Bhushan A, Matherly LH. Impaired membrane transport in methotrexate-resistant CCRF-CEM cells involves early translation termination and increased turnover of a mutant reduced folate carrier. *J Biol Chem* 1999; **274**: 10388-94.

66 Zhao R, Assaraf YG, Goldmann ID. A mutated murine reduced folate carrier (RFC1) with increased affinity for folic acid, decreased affinity for methotrexate, and an obligatory anion requirement for transport function. *J Biol Chem* 1998; **273**: 19065-71.

67 Sadlish H, Williams FMR, Flintoff WF. Functional role of arginine 373 in substrate translocation by the reduce folate carrier. *J Biol Chem* 2002; **277**: 42105-12.

68 Zhao R, Assaraf YG, Goldmann ID. A reduced folate carrier mutation produces substrate-dependent alterations in carrier mobility in murine leukemia cells and methotrexate resistance with conservation of growth in 5-formyltetrahydrofolate. *J Biol Chem* 1998; **273**: 7873-9.

69 Sharina IG, Zhao R, Wang Y, Babani S, Goldman ID. Mutational analysis of the functional role of conserved arginine and lysine residues in transmembrane domains of the murine reduced folate carrier. *Mol Pharmacol* 2001; **59**: 1022-8.

70 Zhao R, Gao F, Wang PJ, Goldmann ID. Role of the amino acid 45 residue in reduced folate carrier function and ion-dependent transport as characterized by site-directed mutagenesis. *Mol Pharmacol* 2000; **57**: 317-23.

71 Tse A, Brigle K, Taylor SM, Moran RG. Mutations in the reduced folate carrier gene which confer dominant resistance to 5,10-dideazatetrahydrofolate. *J Biol Chem* 1998; **273**: 25953-60.

72 Roy K, Tolner B, Chiao JH, Sirotnak FM. A single amino acid difference within the folate transporter encoded by the murine RFC-1 gene selectively alters its interaction with folate analogues. *J Biol Chem* 1998; **273**: 2526-31.

73 Zhao R, Gao F, Goldmann ID. Descrimination among reduced folates and methotrexate as transport substrate by a phenylalanine substitution for serine within the predicted eighth transmembrane domain of the reduced folate carrier. *Biochem Pharmacol* 1999; **58**: 1615-24.

74 Whetstine JR, Gifford AJ, Witt T *et al.* Single nucleotide polymorphism in the human reduced folate carrier: characterization of a high-frequency G/A variant at position 80 and transport properties of the His^{27} and Arg^{27} carriers. *Clin Cancer Res* 2001; **7**: 3416-22.

75 Rots MG, Willey JC, Jansen G *et al.* mRNA expression levels of methotrexate resistance-related proteins in childhood leukemia as determined by a standardized template-based RT-PCR method. *Leukemia* 2000; **14**: 2166-75.

76 Liu S, Song L, Bevins R, Birhiray O, Moscow JA. The murine-reduced folate carrier gene can act as a selectable marker and suicide gene in hematopoietic cells *in vivo*. *Hum Gene Ther* 2002; **13**: 1777-82.